ISSN: 2339-2541
JURNAL GAUSSIAN, Volume 4, Nomor 4, Tahun 2015, Halaman 855-864
Online di: http://ejournal-s1.undip.ac.id/index.php/gaussian

PEMODELAN PERTUMBUHAN EKONOMI JAWA TENGAH MENGGUNAKAN PENDEKATAN *LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR* (LASSO)

Feby Kurniawati Heru Prabowo[1], Yuciana Wilandari[2], Agus Rusgiyono[3]
[1]Mahasiswa Jurusan Statistika FSM Universitas Diponegoro
[2,3]Staff Pengajar Jurusan Statistika FSM Universitas Diponegoro
feby.kurnia.hp@gmail.com, yuciana.wilandari@gmail.com, agus.rusgi@gmail.com

ABSTRACT

The economic growth recently become more important because of its implementation widely, the economic growth concept is a measure of country or regional economy valuation. The economic growth data in this research that is measured by Gross Regional Domestic Product (GRDP) are susceptible of multicollinearity. Multicollinearity become a problem in regression analysis, especially in Ordinary Least Square (OLS) because it causes the regression coefficient estimates become not efficient. One of method to overcome multicollinearity is using Least Absolute Shrinkage and Selection Operator (LASSO). LASSO is a shrinkage method to estimate regression coefficients by minimazing residual sum of squares subject to a constraint. Because of that constraint, LASSO can shrinks coefficients towards zero or set them to exactly zero so it can do variable selection too. Based on Variance Inflation Factor (VIF), there are high correlations between predictor variables, so there is multicollinearity in growth economic data of Jawa Tengah 2013 if we use OLS. In this research, LASSO shrinks eleven coefficients estimator of predictor variables to exactly zero, so that variables considered to have not a significant influence toward model.

Keywords : LASSO, Multicollinearity, Shrinkage, Gross Regional Domestic Product (GRDP)

1. PENDAHULUAN

Melalui proses pembangunan yang terus bergerak dan berjalan, kesempatan-kesempatan bagi masyarakat Indonesia untuk mendapatkan kehidupan yang lebih baik akan menjadi terbuka dan nyata. Salah satu indikator untuk melakukan analisis tentang pembangunan ekonomi suatu negara ataupun daerah adalah dengan melihat pertumbuhan ekonomi daerah tersebut. Dalam penerapannya secara luas, konsep pertumbuhan ekonomi dinilai sebagai tolak ukur penilaian perekonomian suatu negara atau daerah [13].

Pertumbuhan ekonomi dapat diukur dengan peningkatan produksi barang dan jasa atau Pendapatan Nasional. Supaya suatu perekonomian menghasilkan barang dan jasa, diperlukan proses produksi yang membutuhkan sumber daya alam dan diolah dengan menggunakan suatu alat tertentu dan tingkat teknologi tertentu serta sumber daya manusia yang terdidik dan ahli. Oleh karena itu, faktor-faktor yang mampu menjadi dorongan supaya pertumbuhan ekonomi terjadi sangatlah penting untuk diketahui supaya perekonomian tumbuh secara positif dan mantap.

Mengetahui betapa pentingnya pertumbuhan ekonomi bagi suatu negara atau pun daerah, maka penulis melakukan penelitian untuk mengetahui faktor-faktor yang mempengaruhi pertumbuhan ekonomi dengan menggunakan salah satu metode statistika yakni metode LASSO (*Least Absolute Shrinkage and Selection Operator*) dengan studi kasus pada pertumbuhan ekonomi di Jawa Tengah tahun 2013. Metode LASSO cocok diterapkan pada data yang mengandung multikolineritas. Data yang digunakan penulis sangat rentan mengalami masalah multikolinieritas karena antara satu variabel dengan variabel lainnya memiliki hubungan.

Menurut Tibshirani (1996) metode LASSO merupakan metode penyusutan yang melakukan pendugaan koefisien regresi dengan meminimumkan jumlah kuadrat galat dengan suatu kendala $\sum_{j=1}^{k}|\beta_j| \leq t$, dengan t adalah parameter tuning yang mengontrol besarnya penyusutan. Karena kendala tersebut, LASSO mengurangi sejumlah koefisien ke arah nol bahkan tepat nol sehingga dapat melakukan seleksi variabel prediktor. Dengan demikian, model yang dihasilkan dengan metode LASSO menjadi lebih sederhana (parsimoni) dan terhindar dari multikolinieritas.

2. TINJAUAN PUSTAKA

2.1 Konsep Dasar Pertumbuhan Ekonomi

Tingkat pertumbuhan ekonomi menggambarkan tentang kenaikan riil dari produksi barang dan jasa yang dihasilkan oleh suatu negara dalam suatu tahun tertentu. Pembangunan ekonomi adalah pertumbuhan ekonomi ditambah dengan perubahan [10].

Terdapat beberapa macam alat yang dapat digunakan untuk mengukur pertumbuhan ekonomi yaitu Produk Domestik Bruto (PDB) / Produk Domestik Regional Bruto (PDRB) maupun PDB Per Kapita/PDRB Per Kapita [11].

Beberapa teori yang membahas mengenai pertumbuhan ekonomi yakni [10]:

1. Teori pertumbuhan klasik, menekankan pada peranan sumber daya alam, modal dan tenaga kerja dalam pertumbuhan ekonomi;
2. Teori Harrod-Domar, memberikan peranan kunci kepada investasi di dalam proses pertumbuhan ekonomi;
3. Teori pertumbuhan neoklasik, mengemukakan bahwa laju pertumbuhan ekonomi suatu negara dipengaruhi oleh perkembangan teknologi, pertambahan stok modal dan pertambahan tenaga kerja; dan
4. Teori pertumbuhan Schumpeter, mengemukakan bahwa pengusaha yang kreatif dan inovatif merupakan tokoh kunci dalam mengenalkan produk baru dan perbaikan terus menerus sehingga membawa pada pertumbuhan ekonomi.

Terdapat lima (5) faktor yang mempengaruhi pertumbuhan ekonomi, yakni modal atau kapital, sumber daya alam, *human capital*, angkatan kerja yang bekerja, kemajuan teknologi dan wirausahawan [7].

2.2 Metode Kuadrat Terkecil (*Ordinary Least Square*/OLS)

Suatu model regresi linier menyatakan hubungan antara satu variabel respon dengan satu atau lebih variabel prediktor yang memiliki hubungan yang linier dalam parameter. Adapun model pengamatan ke-i jika ada sebanyak n pengamatan sebagai berikut [9]:

$$\mathrm{Y_i} = \beta_0^0 + \sum_{j=1}^{k} \beta_j^0 \mathrm{X_{ij}} + \varepsilon_i \text{ dengan } i = 1,2,\ldots,n \quad (1)$$

Dalam notasi matriks, Persamaan (1) dapat ditulis menjadi persamaan berikut:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}^{\mathbf{0}} + \boldsymbol{\varepsilon}$$

dengan

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ . \\ . \\ Y_n \end{pmatrix}, \mathbf{X} = \begin{pmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nk} \end{pmatrix}, \boldsymbol{\beta}^{\mathbf{0}} = \begin{pmatrix} \beta_0^0 \\ \beta_1^0 \\ \vdots \\ \beta_k^0 \end{pmatrix} \text{ dan } \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Tujuan dari metode OLS adalah meminimumkan jumlah kuadrat galat (JKG), yakni dengan meminimumkan persamaan berikut:

$$\text{JKG} = \sum_{i=1}^{n} \left(\mathrm{Y_i} - \hat{\mathrm{Y}}_i\right)^2 = \sum_{i=1}^{n} \varepsilon_i^2 = \boldsymbol{\varepsilon}^{\mathbf{T}}\boldsymbol{\varepsilon}$$

Sehingga solusi untuk estimasi parameter regresi adalah:

$$\hat{\boldsymbol{\beta}}^{\mathbf{0}} = (\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}(\mathbf{X}^{\mathbf{T}}\mathbf{Y})$$

Untuk mengetahui ketepatan model regresi dapat dilihat dari uji signifikansi regresi dan uji signifikansi parameter individual. Uji signifikansi regresi atau uji F dimaksudkan untuk menentukan apakah terdapat hubungan linier antara variabel respon Y dengan himpunan variabel prediktor $\{X_1, X_2, \ldots, X_k\}$ [9]. Model regresi signifikan apabila nilai $F_{hitung} \geq F_{\alpha(k,n-p)}$ atau jika menggunakan software *p-value* $\leq \alpha$.

Uji signifikansi parameter individual atau uji t digunakan untuk mengetahui variabel prediktor mana yang pengaruhnya tidak nyata terhadap variabel respon [11]. Koefisien parameter β_j^0 signifikan terhadap model apabila nilai $|t_{hitung}| \geq t_{\frac{\alpha}{2},n-p}$ atau *p-value* $\leq \alpha$.

Nilai koefisien determinasi pada intinya mengukur seberapa jauh kemampuan model dalam menerangkan variasi variabel respon [4]. Nilai R^2 dapat dihitung menggunakan rumus berikut [11]:

$$R^2 = \frac{JKR}{JKT} = 1 - \frac{JKG}{JKT}$$

dengan JKT (Jumlah Kuadrat Total) = JKR + JKG

Misalkan terdapat data berpasangan (X,Y), koefisien korelasi (r_{XY}) menyatakan derajat hubungan linier antara X dan Y tanpa mempersoalkan hubungan kausal (sebab-akibat) [14]. Koefisien korelasi memiliki nilai $-1 \leq r_{XY} \leq 1$. Apabila r bernilai positif maka hubungan tersebut searah dan apabila r bernilai negatif maka hubungan tersebut berlawanan arah. Semakin besar nilai r_{XY} maka hubungan linier antara X dan Y semakin tinggi atau kuat.

2.3 Multikolinieritas

Multikolinieritas adalah terjadinya hubungan linier antar variabel prediktor dalam suatu model regresi linier berganda [5]. Dampak adanya multikolinieritas dalam model regresi linier berganda adalah penaksir mempunyai variansi dan kovariansi yang besar sehingga membuat variabel prediktor secara statistik tidak signifikan mempengaruhi variabel respon. Salah satu cara untuk mendeteksi adanya multikolinieritas adalah dengan memperhatikan nilai *Variance Inflation Factor* (VIF). Nilai VIF untuk variabel prediktor ke-j dirumuskan sebagai berikut:

$$VIF_j = (1 - R_j^2)^{-1}$$

Jika nilai VIF lebih besar dari 10, maka menunjukkan adanya multikolinieritas.

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

Tibshirani memperkenalkan metode baru untuk meningkatkan estimasi OLS yakni LASSO (*Least Absolute Shrinkage and Selection Operator*). Penduga koefisien pada LASSO diperoleh dengan cara meminimumkan persamaan berikut [12]:

$$JKG = \sum_{i=1}^{n}\left(Y_i^* - \beta_0 - \sum_{j=1}^{k}\beta_j X_{ij}^*\right)^2 \quad \text{dengan syarat } \sum_{j=1}^{k}\left|\beta_j\right| \leq t$$

Jika $\hat{\beta}_j^0$ merupakan penduga kuadrat terkecil dan $t_0 = \sum_{j=1}^{k}\left|\hat{\beta}_j^0\right|$, maka apabila $t < t_0$ akan menyebabkan koefisien penduga kuadrat terkecil menyusut ke arah nol dan memungkinkan beberapa koefisien tepat nol. Sehingga LASSO dapat melakukan peran sebagai *variable selection* sekaligus mengatasi multikolinieritas.

Penduga koefisien LASSO diperoleh dengan menentukan parameter tuning yang dibakukan, yaitu $s = \frac{t}{\sum_{j=1}^{k}|\hat{\beta}_j^0|}$ dengan $t = \sum_{j=1}^{k}\left|\hat{\beta}_j\right|$ dan $\hat{\beta}_j^0$ adalah penduga kuadrat terkecil (OLS) atau pada output plot LASSO s ditulis sebagai *|beta|/max|beta|*. Nilai optimal s dapat diperoleh melalui *cross validation* (CV) [2].

2.5 Algoritma LARS

Least Angle Regression (LARS) merupakan metode seleksi model dimana algoritmanya dapat dimodifikasi untuk diimplementasikan ke dalam penyelesaian LASSO. Menurut Hastie *et al.* (2008) algoritma LARS adalah sebagai berikut:

1. Membakukan variabel prediktor dan variabel respon sehingga memiliki nilai tengah nol dan ragam satu. Mulai dengan galat $\mathbf{e} = \mathbf{Y}^* - \overline{\mathbf{Y}}^*$ dan $\beta_1, \beta_2, \ldots, \beta_k = 0$.
2. Mencari variabel prediktor $\mathbf{X}_\mathbf{q}^*$ yang paling berkorelasi dengan $\mathbf{e}$.
3. Mengubah nilai β_q dari 0 bergerak ke arah koefisien *least-squares* $\mathbf{e}$ dengan $\mathbf{X}_\mathbf{q}^*$ sampai variabel prediktor yang lain $\mathbf{X}_\mathbf{r}^*$ memiliki korelasi yang sama dengan galat sekarang akibat $\mathbf{X}_\mathbf{q}^*$.
4. Mengubah β_q dan β_r bergerak menuju arah yang ditentukan oleh koefisien *joint least-squares* $\mathbf{e}$ dengan $(\mathbf{X}_\mathbf{q}^*, \mathbf{X}_\mathbf{r}^*)$ sampai variabel prediktor yang lain $\mathbf{X}_\mathbf{s}^*$ memiliki korelasi yang sama dengan galat sekarang akibat $(\mathbf{X}_\mathbf{q}^*, \mathbf{X}_\mathbf{r}^*)$.
5. Mengulang sesuai langkah ke-4 sampai semua k prediktor dimasukkan ke dalam model. Setelah min(n-1,k) langkah, solusi model penuh *least-squares* didapatkan.

Modifikasi algoritma LARS untuk mendapatkan solusi LASSO adalah dengan mengubah langkah ke-4 menjadi berikut:

4a. Apabila koefisien variabel yang bukan nol mencapai nol, variabel tersebut dikeluarkan dari gugus variabel aktif dan dihitung kembali arah dari *joint least-squares* sekarang.

LARS selalu mengambil k langkah untuk mendapatkan penduga kuadrat terkecil secara penuh, sedangkan modifikasi LARS untuk LASSO dapat memiliki lebih dari k langkah untuk mendapatkannya.

2.6 Penghitungan Penduga Koefisien LASSO dengan Algoritma LARS

Didefinisikan $\widehat{\boldsymbol{\mu}}_\mathbf{0} = \mathbf{X}^*\widehat{\boldsymbol{\beta}} = 0$. Nilai $\widehat{\boldsymbol{\mu}}$ akan berubah seiring tahapan berjalan. Misalkan $\widehat{\boldsymbol{\mu}}_{\boldsymbol{A}}$ adalah nilai estimasi dengan variabel aktif A [3].

1. Menghitung vektor korelasi $\hat{\mathbf{c}} = {\mathbf{X}^*}^{\mathbf{T}}(\mathbf{Y}^* - \widehat{\boldsymbol{\mu}}_{\boldsymbol{A}})$ dan nilai korelasi absolut terbesar $\hat{\mathrm{C}} = \underbrace{\max}_{\mathrm{j}} \{|\hat{\mathrm{c}}_\mathrm{j}|\}$ sehingga $A = \{j \mid |\hat{\mathrm{c}}_\mathrm{j}| = \hat{\mathrm{C}}\}$
2. Menghitung *equiangular vector* $(\mathbf{u}_A)$. Didefinisikan $\mathbf{X}_A = (\ldots \mathrm{s_j}\boldsymbol{X}_{\boldsymbol{j}}^* \ldots)_{j\in A}$ dengan $\mathrm{s_j} = sign\{\hat{\mathrm{c}}_\mathrm{j}\}$ untuk $\mathrm{j} \in A$ dan $\boldsymbol{\omega}_A = \mathrm{A}_A\mathbf{G}_A^{-\mathbf{1}}\mathbf{1}_A$ dimana $\mathrm{A}_A = \left(\mathbf{1}_A^{\boldsymbol{T}}\mathbf{G}_A^{-\mathbf{1}}\mathbf{1}_A\right)^{-\frac{1}{2}}$. Nilai *equiangular vector* didapat dari $\mathbf{u}_A = \mathbf{X}_A\boldsymbol{\omega}_A$
3. Menghitung panjang dari $\mathbf{u}_A$ $(\hat{\gamma})$. Didefinisikan vektor *inner product* $\mathbf{a} \equiv {\mathbf{X}^*}^{\mathbf{T}}\mathbf{u}_{\boldsymbol{A}}$, sehingga $\hat{\gamma}$ dapat diperoleh dengan persamaan berikut: $\hat{\gamma} = min^{+}_{j\in A^c}\left\{\frac{\hat{C}-\hat{c}_j}{\mathrm{A}_A - a_j}, \frac{\hat{C}+\hat{c}_j}{\mathrm{A}_A + a_j}\right\}$
4. Langkah selanjutnya adalah memperbarui nilai $\widehat{\boldsymbol{\mu}}_A$, yakni $\widehat{\boldsymbol{\mu}}_{\boldsymbol{A}+} = \widehat{\boldsymbol{\mu}}_A + \hat{\gamma}\mathbf{u}_A$
5. Algoritma LARS untuk LASSO mengharuskan $sign(\hat{\beta}_\mathrm{j}) = sign(\hat{\mathrm{c}}_\mathrm{j}) = \mathrm{s_j}$. Kondisi ini dapat diperoleh apabila $\tilde{\gamma} = \underbrace{\min}_{\gamma_\mathrm{j}>0} \{\gamma_\mathrm{j}\}$ dengan $\gamma_\mathrm{j} = -\hat{\beta}_\mathrm{j}/(\mathrm{s_j}\boldsymbol{\omega}_{A_\mathrm{j}})$. Apabila $\tilde{\gamma}$ kurang dari $\hat{\gamma}$ maka $\beta_\mathrm{j}(\gamma)$ 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 penghitungan arah *equiangular* selanjutnya. $\widehat{\boldsymbol{\mu}}_{A+} = \widehat{\boldsymbol{\mu}}_A + \tilde{\gamma}\mathbf{u}_{\boldsymbol{A}}$ dan $A_+ = A - \{j\}$.
   Variabel j dimasukkan kembali ke dalam penghitungan LARS pada tahap selanjutnya.

2.7 Validasi Silang

Validasi silang membagi data menjadi dua bagian, yakni data *training* dan data *testing*. Data *training* digunakan untuk menentukan nilai $\hat{\boldsymbol{\beta}}$ atau untuk menyusun model, sedangkan data *testing* digunakan untuk menguji kebaikan $\mathbf{X}\hat{\boldsymbol{\beta}}$. Nilai validasi silang yang diperoleh merupakan penduga bagi galat prediksi [6]. Salah satu metode tipe validasi silang adalah *k-fold*. Nilai galat prediksi *k-fold* CV apabila data dipartisi menjadi c bagian diperoleh dengan persamaan berikut [8]:

$$\text{CV MSE} = \frac{1}{c}\sum_{h=1}^{c} MSE_h = \frac{1}{c}\sum_{h=1}^{c}\sum_{(x_i,y_i)\in S}\left(y_i - \hat{y}_{-c}(x_i)\right)^2$$

dengan $\hat{y}_{-c}(x_i)$ adalah dugaan Y untuk X_i pada data testing S pada saat *fold* ke-c tidak digunakan dalam menduga model dan Y_i adalah variabel respon ke-i pada data testing S. Validasi silang yang sebaiknya digunakan adalah validasi silang *5-fold* atau *10-fold* karena menghasilkan nilai CV dengan bias tinggi tetapi ragam rendah [8].

3. METODE PENELITIAN

Data yang digunakan adalah data sekunder yang bersumber dari publikasi online Badan Pusat Statistik (BPS) Provinsi Jawa Tengah. Publikasi online tersebut dapat diperoleh melalui website www.jateng.bps.go.id. Variabel penelitian yang diteliti berupa variabel respon (Y), yakni Produk Domestik Regional Bruto (PDRB) Jawa Tengah atas dasar harga konstan 2000 tahun 2013 dan variabel prediktor yang meliputi:

Tabel 1. Variabel Prediktor yang Digunakan

Faktor	Variabel prediktor
Modal atau Kapital	X_1 = Posisi tabungan (juta rupiah) X_2 = Realisasi pajak daerah (juta rupiah)
Sumber daya alam	X_3 = Luas hutan (ha) X_4 = Produktivitas padi (ku/ha)
Human Capital	X_5 = Angka Kesakitan (persen) X_6 = Angka Harapan Hidup (usia) X_7 = Rata-rata lama sekolah (tahun) X_8 = Angka Melek Huruf (persen) X_9 = Persentase pengeluaran per kapita per bulan untuk makanan (persen) X_{10} = Kebutuhan hidup layak (juta rupiah)
Angkatan kerja yang bekerja	X_{11} = Jumlah angkatan kerja yang bekerja berdasarkan pendidikan tertinggi yang ditamatkan (SLTP dan SLTA+) (jiwa)
Kemajuan Teknologi	X_{12} = Persentase rumah tangga yang memiliki telepon seluler (persen) X_{13} = Persentase rumah tangga yang memiliki komputer (persen)
Wirausahawan	X_{14} = Jumlah wirausahawan (jiwa)

4. HASIL DAN PEMBAHASAN

4.1 Gambaran Umum Data Penelitian

Untuk mengetahui seberapa besar hubungan dari masing-masing variabel data yang digunakan dapat diketahui dengan melihat nilai koefisien korelasi. Hasil penghitungan nilai koefisien korelasi untuk setiap variabel data yang digunakan dapat dilihat pada Tabel 2.

Terlihat pada Tabel 2 bahwa terdapat lebih dari satu variabel prediktor yang memiliki korelasi tinggi dengan variabel prediktor lainnya. Sehingga hal ini menjadi indikasi awal bahwa terdapat masalah multikolinieritas pada data yang digunakan.

Tabel 2. Nilai Koefisien Korelasi Pearson Antar Variabel Penelitian

	Y	**X_1**	**X_2**	**X_3**	**X_4**	**X_5**	**X_6**	**X_7**	**X_8**	**X_9**	**X_{10}**	**X_{11}**	**X_{12}**	**X_{13}**	**X_{14}**
Y	1,00	0,78	0,83	0,01	-0,23	0,16	0,14	0,30	0,19	-0,35	0,68	0,75	0,32	0,37	0,23
X_1	0,78	1,00	0,95	-0,22	-0,16	-0,07	0,29	0,58	0,39	-0,63	0,61	0,64	0,50	0,69	-0,03
X_2	0,83	0,95	1,00	-0,19	-0,21	-0,06	0,24	0,50	0,36	-0,56	0,69	0,71	0,44	0,61	0,03
X_3	0,01	-0,22	-0,19	1,00	-0,24	0,29	-0,09	-0,67	-0,53	0,27	-0,10	0,04	-0,55	-0,56	0,65
X_4	-0,23	-0,16	-0,21	-0,24	1,00	-0,33	0,06	0,26	0,03	-0,15	-0,39	-0,13	0,11	0,21	-0,14
X_5	0,16	-0,07	-0,06	0,29	-0,33	1,00	-0,39	-0,37	-0,14	0,11	-0,09	0,26	-0,38	-0,37	0,41
X_6	0,14	0,29	0,24	-0,09	0,06	-0,39	1,00	0,32	-0,01	0,41	0,22	0,29	0,22	0,29	-0,09
X_7	0,30	0,58	0,50	-0,67	0,26	-0,37	0,32	1,00	0,67	-0,71	0,34	0,15	0,85	0,93	-0,61
X_8	0,19	0,39	0,36	-0,53	0,03	-0,14	-0,01	0,67	1,00	-0,27	0,42	-0,06	0,63	0,56	-0,47
X_9	-0,35	-0,63	-0,56	0,27	-0,15	0,11	0,41	-0,71	-0,27	1,00	-0,29	-0,35	-0,56	-0,76	0,29
X_{10}	0,68	0,61	0,69	-0,10	-0,39	-0,09	0,22	0,34	0,42	-0,29	1,00	0,37	0,43	0,35	-0,12
X_{11}	0,75	0,64	0,71	0,04	-0,13	0,26	0,29	0,15	-0,06	-0,35	0,37	1,00	0,01	0,17	0,49
X_{12}	0,32	0,50	0,44	-0,55	0,11	-0,38	0,22	0,85	0,63	-0,56	0,43	0,01	1,00	0,84	-0,60
X_{13}	0,37	0,69	0,61	-0,56	0,21	-0,37	0,29	0,93	0,56	-0,76	0,35	0,17	0,84	1,00	-0,52
X_{14}	0,23	-0,03	0,03	0,65	-0,14	0,41	-0,09	-0,61	-0,47	0,29	-0,12	0,49	-0,60	-0,52	1,00

Data dari variabel penelitian perlu dikonversi ke dalam bentuk normal standar. Sebelum melakukan penstandaran data perlu dilakukan pengujian normal multivariat terhadap data penelitian yang digunakan.

Dari hasil pengujian normal multivariate, jika diambil tingkat kesalahan 5%, didapat *p-value* = 0,9794. Maka didapat kesimpulan bahwa data penelitian berdistribusi normal multivariat.

4.2 Pendugaan Koefisien Menggunakan OLS

Hasil pengolahan data dengan menggunakan software Minitab 14 diperoleh persamaan regresi sebagai berikut :

$$\hat{Y}^* = 0{,}3097\, X_1^* + 0{,}1299\, X_2^* + 0{,}0447\, X_3^* + 0{,}0933\, X_4^* + 0{,}0662\, X_5^* - 0{,}1570\, X_6^* - 0{,}0345\, X_7^* - 0{,}1171\, X_8^* - 0{,}1441\, X_9^* + 0{,}3482\, X_{10}^* + 0{,}3254\, X_{11}^* + 0{,}3352\, X_{12}^* - 0{,}2774\, X_{13}^* + 0{,}0810\, X_{14}^*$$

Hasil pengolahan data menggunakan Minitab 14 didapat nilai F_{hitung} sebesar 8,52 dan nilai *p-value* sebesar 0,000. Nilai $F_{\alpha(k,n-p)}$ dengan α = 5%, k = 14 dan n - p = 20 adalah 2,22. Karena nilai $F_{hitung} \geq F_{tabel}$ (8,52 ≥ 2,22) atau *p-value* ≤ α (0,000 ≤ 0,05), maka didapat kesimpulan bahwa model regresi yang diuji sudah signifikan atau terdapat pengaruh secara bersama-sama dari variabel prediktor terhadap variabel respon.

Nilai t_{tabel} α = 5% dengan derajat bebasnya adalah 20 adalah 2,086. Hasil pengolahan data menggunakan Minitab 14, didapat nilai t_{hitung} dan *p-value* berikut:

Tabel 3. Hasil Pengujian Statistik t

Variabel Prediktor	t_{hitung}	*p-value*
X_1	0,89	0,385
X_2	0,27	0,787
X_3	0,29	0,772
X_4	0,80	0,434
X_5	0,47	0,646
X_6	-1,27	0,219
X_7	-0,08	0,938
X_8	-0,65	0,526
X_9	-0,83	0,415
X_{10}	2,16	0,043
X_{11}	1,08	0,292
X_{12}	1,72	0,101
X_{13}	-0,66	0,519
X_{14}	0,38	0,708

Jadi, pada taraf signifikansi 5% didapat hasil bahwa variabel prediktor yang memiliki pengaruh signifikan terhadap variabel respon hanya variabel X_{10} yakni variabel kehidupan hidup layak.

Hasil pengolahan data menggunakan Minitab 14 didapat nilai R^2 pada model regresi ini sebesar 0,856. Hal ini berarti bahwa besarnya pengaruh keempat belas variabel prediktor terhadap variabel respon adalah sebesar 85,6%, sedangkan sisanya yakni 14,4% dipengaruhi faktor-faktor lain di luar model.

4.3 Pendeteksian Multikolinieritas

Berikut adalah nilai VIF masing-masing variabel independen hasil pengolahan software Minitab 14.

Tabel 4. Nilai VIF untuk Setiap Variabel Prediktor

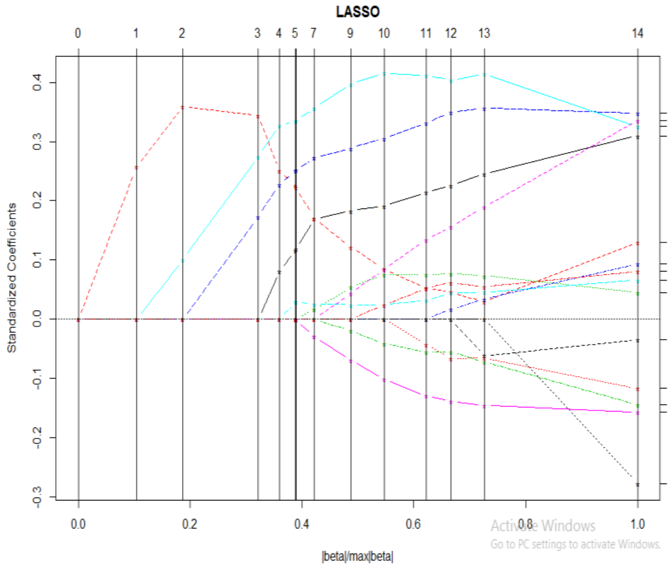
Variabel Prediktor	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}
Nilai VIF	16,9	31,4	3,2	1,9	2,8	2,1	26,6	4,6	4,2	3,6	12,6	5,3	24,8	6,3

Berdasarkan Tabel 4 diketahui bahwa terdapat lima variabel prediktor yang memiliki niai VIF lebih besar dari 10, sehingga dapat disimpulkan bahwa terdapat masalah multikolinieritas pada model regresi ini. Selain itu multikolinieritas juga dapat dideteksi dengan melihat nilai koefisien determinasinya. Pada model regresi ini, koefisien determinasinya bernilai cukup besar, yakni 85,6% tetapi banyak variabel prediktor yang tidak signifikan berpengaruh terhadap variabel respon.

4.4 Pendugaan Koefisien Menggunakan LASSO

4.4.1 Tahapan-Tahapan Variabel Prediktor yang Masuk ke Model LASSO

Pendugaan koefisien LASSO dilakukan secara bertahap dengan menetapkan koefisien awal semuanya bernilai 0. Selanjutnya secara bertahap variabel prediktor yang paling berkorelasi dengan galat akan masuk ke dalam model. Berikut adalah plot tahapan variabel prediktor yang masuk ke dalam model:



Gambar 1. Plot Obyek LARS yang dihasilkan oleh fungsi LARS untuk Menduga Koefisien LASSO

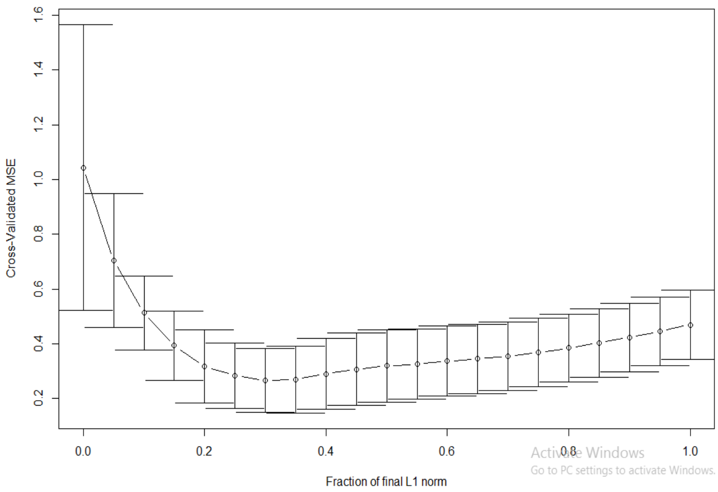
Variabel X_2 adalah variabel prediktor pertama yang masuk ke dalam model karena memiliki korelasi tertinggi dengan galat. Nilai korelasi untuk variabel X_2 adalah c_2 = 28,994. Kemudian penduga koefisien parameter dari X_2 bergerak seiring dengan pergerakan nilai s, yakni dari s = 0 sampai s = 1. Pada tahap kedua variabel X_{11} masuk ke dalam model ketika nilai s = 0,10460. Notasi s disini adalah parameter tuning yang dibakukan. Variabel X_{11} adalah variabel prediktor kedua yang masuk ke dalam model karena variabel prediktor ini memiliki korelasi yang sama dengan galat yang dihasilkan dari X_2. Jadi pada tahap ini c_2 = c_{11} = 19,975. Proses ini terus berlanjut sampai semua variabel prediktor masuk ke model. Berikut adalah tahapan variabel prediktor yang masuk ke model LASSO:

Tabel 5. Variabel prediktor yang masuk ke model untuk setiap tahapan pada metode LASSO

Tahap	Variabel Prediktor yang Masuk ke Model	Tahap	Variabel Prediktor yang Masuk ke Model
1	X_2	8	X_9
2	X_{11}	9	X_{12}
3	X_{10}	10	X_{14}
4	X_1	11	X_8
5	X_5	12	X_4
6	X_6	13	X_7
7	X_3	14	X_{13}

4.4.2 Pemilihan Model Terbaik LASSO

Pemilihan model terbaik LASSO ditentukan berdasarkan nilai s yang meminimumkan CV MSE. Pemilihan ini dilakukan dengan menggunakan *k-fold Cross Validation* (*k-fold* CV). Nilai s menunjukkan $\frac{\sum|\hat{\beta}_j|}{\sum|\hat{\beta}_j^0|}$ dimana $\sum|\hat{\beta}_j|$ adalah jumlahan absolut dari penduga koefisien LASSO sedangkan $\sum|\hat{\beta}_j^0|$ adalah jumlahan absolut dari penduga koefisien OLS.



Gambar 2. Plot MSE Menggunakan CV Pada Variasi Nilai s

Berdasarkan plot MSE untuk setiap variasi nilai s di atas titik minimum plot berada pada nilai s = 0,30. Nilai CV MSE yang paling minimum tersebut dapat berbeda setiap kali melakukan pemanggilan fungsinya. Dari beberapa pengulangan diperoleh nilai CV berubah-ubah pada nilai sekitar 0,30. Pada nilai s = 0,30 terdapat tiga variabel prediktor yang masuk ke dalam model. Variabel prediktor yang tidak masuk ke dalam model berarti memiliki penduga koefisien parameter bernilai nol dan terseleksi dari model. Berikut adalah persamaan regresi yang dihasilkan dengan menggunakan metode LASSO:

$$\widehat{Y}^* = 0,3466436\ X_2^* + 0,1455690\ X_{10}^* + 0,2469213\ X_{11}^*$$

4.5 Perbandingan Nilai Penduga Koefisien Parameter Metode OLS dengan Metode LASSO

Batasan *shrinkage* pada metode LASSO menyebabkan nilai penduga koefisien parameter menyusut sehingga variabel prediktor yang penting atau berpengaruh terhadap model tetap dimasukkan ke dalam model, sedangkan variabel prediktor yang kurang penting akan disusutkan sampai nol dan terseleksi dari model sehingga model menjadi lebih efisien. Perbandingan nilai penduga koefisien parameter dengan metode OLS dan metode LASSO dapat dilihat pada Tabel 6.

Tabel 6. Nilai Penduga Koefisien Parameter Metode OLS dan Metode LASSO

Variabel Prediktor	OLS	LASSO
X_1	0,3097	0,0000
X_2	0,1299	0,3466
X_3	0,0447	0,0000
X_4	0,0933	0,0000
X_5	0,0662	0,0000
X_6	-0,1570	0,0000
X_7	-0,0345	0,0000
X_8	-0,1171	0,0000
X_9	-0,1441	0,0000
X_{10}	0,3482	0,1456
X_{11}	0,3254	0,2469
X_{12}	0,3352	0,0000
X_{13}	-0,2774	0,0000
X_{14}	0,0810	0,0000

Pada Tabel 6 dapat diketahui bahwa nilai penduga koefisien parameter dengan metode LASSO cenderung menyusut ke arah nol atau cenderung memiliki pengaruh lebih kecil terhadap variabel respon daripada nilai penduga koefisien parameter dengan metode OLS. Nilai penduga koefisien parameter untuk variabel X_1, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{12}, X_{13} dan X_{14} disusutkan sampai tepat nol, sehingga variabel-variabel tersebut tidak memiliki pengaruh atau kurang penting terhadap model. Melalui penyusutan yang sampai tepat nol ini, LASSO dapat pula digunakan sebagai metode seleksi variabel. Melalui seleksi variabel ini, model menjadi lebih sederhana dan efisien serta dapat mengatasi masalah multikolinieritas.

5. KESIMPULAN

Berdasarkan analisis dan pembahasan diperoleh kesimpulan sebagai berikut:

1. Pemilihan model terbaik LASSO menggunakan *Cross-Validation* dan didapat model terbaik pada s = 0.30 dengan modelnya adalah sebagai berikut :

$$\hat{Y}^* = 0{,}3466436\, X_2^* + 0{,}1455690\, X_{10}^* + 0{,}2469213\, X_{11}^*$$

2. Pada model regresi dengan metode LASSO, variabel prediktor yang signifikan mempengaruhi model adalah Realisasi Pajak (X_2), Kebutuhan Hidup Layak (X_{10}) dan Jumlah Angkatan Kerja yang Bekerja (X_{11}).

6. DAFTAR PUSTAKA

[1] BPS Provinsi Jawa Tengah. www.jateng.bps.go.id (Diakses 23 Maret 2015)

[2] Dewi, Y.S. 2010. OLS, LASSO dan PLS pada Data Mengandung Multikolinieritas. *Jurnal Ilmu Dasar* Vol. 11, No. 1: Hal. 83-91

[3] Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. 2004. Least Angle Regression. *The Annals of Statistics* Vol. 32, No. 2 : Hal. 407-499

[4] Ghozali, I. 2011. *Aplikasi Analisis Multivariat dengan Program IBM SPSS 19*. Semarang: Badan Penerbit Universitas Diponegoro

[5] Gujarati, D. 2003. *Ekonomi Dasar*. Jakarta: Erlangga

[6] Hastie, T., Tibshirani, R., Friedman, J. 2008. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Edisi Kedua. Springer : New York

[7] Irawan dan Suparmoko, M. 2002. *Ekonomika Pembangunan*. Edisi Keenam. BPFE:Yogyakarta

[8] James, G., Witten D., Hastie, T., Tibshirani, R. 2013. *An Introduction to Statistical Learning with Applications in R*. Springer : New York

[9] Montgomery, D.C. and Runger, G.C. 2011. *Applied Statistics and Probability for Engineers*. John Wiley & Sons : New York

[10] Sukirno, S. 2006. *Ekonomi Pembangunan : Proses, Masalah, dan Dasar Kebijakan*. Edisi kedua. Jakarta: Kencana

[11] Suparmoko, M. 2000. *Pengantar Ekonomika Makro*. Edisi Keempat. Yogyakarta: BPFE

[12] Tibshirani, R. 1996. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society Series B (Methodological)* Vol. 58, No. 1 : Hal. 267-288

[13] Todaro, M.P. 1994. *Pembangunan Ekonomi*. Edisi Kelima. Haris Munandar, penerjemah. Jakarta: Bumi Aksara. Terjemahan dari: Economic Development

[14] Usman, H., Akbar, R.P.S. 2008. *Pengantar Statistika*. Edisi Kedua. Jakarta: Bumi Aksara