Buletin Ilmiah Math. Stat dan Terapannya (Bimaster)
Volume 11, No.1 (2022), hal 111-120

PEMODELAN PERTUMBUHAN EKONOMI KALIMANTAN BARAT MENGGUNAKAN PENDEKATAN *LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR* (LASSO)

Oktavianus Frans L, Setyo Wira Rizki, Dadan Kusnandar

INTISARI

Pertumbuhan ekonomi suatu daerah dapat diukur dengan peningkatan produksi barang dan jasa serta pendapatan nasional. Salah satu cara untuk menentukan faktor-faktor yang mempengaruhi pertumbuhan ekonomi yaitu dengan menggunakan analisis regresi. Dalam analisis regresi ada beberapa pelanggaran yang sering terjadi terhadap asumsi-asumsinya, salah satunya yaitu terjadinya multikolinearitas. Dalam mengatasi masalah multikolinearitas yang terjadi terdapat beberapa metode yang dapat digunakan salah satunya ialah metode least absolute shrinkage and selection operator (LASSO). Pada penelitian ini, variabel respon yang digunakan yaitu produk domestik regional bruto (PDRB). Sedangkan variabel prediktor diantaranya penanaman modal asing (PMA), angkatan kerja (AK), penanaman modal dalam negeri (PMDN) serta dana alokasi umum (DAU). Berdasarkan uji asumsi klasik pada analisis regresi terjadi masalah multikolinearitas pada variabel PMDN karena memiliki nilai VIF >10. Sehingga dilakukan penanganan dengan menggunakan metode LASSO. Dengan menggunakan metode LASSO, nilai penduga koefisien parameter untuk variabel PMDN disusutkan sampai tepat nol, sehingga variabel PMDN tidak memiliki pengaruh terhadap model. Berdasarkan perbandingan nilai R-Square, metode LASSO lebih tinggi yaitu sebesar 98,70%, dibandingkan dengan metode kuadrat terkecil pada analisis regresi sebesar 95.54%. Sehingga metode LASSO dapat digunakan sebagai metode seleksi variabel sehingga model menjadi lebih sederhana dan efisien serta dapat mengatasi masalah multikolinieritas.

Kata kunci: *PDRB, Regresi, Multikolinearitas*

PENDAHULUAN

Pertumbuhan ekonomi di suatu daerah dapat diukur tidak hanya dari pendapatan nasional, tetapi juga dalam produksi barang dan jasa.. 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 mempengaruhi pertumbuhan ekonomi sangatlah penting untuk diketahui agar perekonomian tumbuh secara positif dan mantap. Salah satu cara untuk menentukan faktor-faktor yang mempengaruhi pertumbuhan ekonomi yaitu dengan menggunakan analisis regresi. Pada umumnya analisis regresi mempelajari pola dan mengukur hubungan statistik antara dua atau lebih variabel yang bertujuan untuk mendapatkan model regresi. Model regresi adalah suatu cara formal untuk mengekspresikan dua unsur penting suatu hubungan statistik, yaitu kecenderungan berubahnya variabel terikat secara sistematis sejalan dengan berubahnya variabel bebas dan berpencarnya titik-titik disekitar kurva taksiran model[1]. Salah satu metode pendugaan parameter model regresi linear berganda adalah metode kuadrat terkecil (MKT).

Dalam analisis regresi ada beberapa pelanggaran yang sering terjadi terhadap asumsi-asumsinya, salah satunya yaitu terjadinya multikolinearitas. Adanya multikolinearitas dalam model regresi linear berganda menyebabkan penduga memiliki varians dan kovarians yang besar. Dengan demikian interval kepercayaan untuk paramer regresi cenderung sangat lebar dan diperoleh hasil yang tidak signifikan secara statistik. Dalam penelitian ini, model regresi yang dibentuk dari faktor-faktor yang mempengaruhi pertumbuhan ekonomi di wilayah Kalimantan Barat pada tahun 2017 berdasarkan Kabupaten/Kota sangat rentan mengalami masalah multikolinearitas sehingga perlu untuk diatasi.

Dalam mengatasi masalah multikolinearitas yang terjadi terdapat beberapa metode yang dapat digunakan salah satunya ialah metode *least absolute shrinkage and selection operator* (LASSO).

Metode LASSO pertama kali dikenalkan oleh Tibshirani tahun 1996 dan merupakan teknik regresi yang melakukan pendugaan dengan meminimumkan jumlah kuadrat galat, dimana koefisien regresi lebih kecil dari parameter *tuning*. Parameter *tuning* adalah parameter yang mengontrol besarnya penyusutan koefisien regresi menuju nol. Karena kendala tersebut, LASSO dapat menyusutkan beberapa koefisien regresi bahkan menjadi mendekati nol atau tepat nol. Sehingga LASSO juga dapat berfungsi sebagai seleksi variabel yang mampu mengatasi masalah multikolinearitas. Dalam penyelesaian LASSO diperlukan algoritma LARS (*least angle regression*) untuk melakukan seleksi variabel. Modifikasi LARS merupakan algoritma yang efisien dalam memperoleh penduga koefisien LASSO.

ANALISIS REGRESI

Analisis regresi adalah salah satu metode statistika yang digunakan untuk menyelidiki atau membangun model hubungan antara dua variabel atau lebih. Apabila variabel-variabel regresi tersebut berhubungan secara linier, maka disebut dengan regresi linier. Regresi linier yang menghubungkan antara satu variabel respon dan satu variabel prediktor disebut regresi linier sederhana. Sedangkan regresi linier yang menghubungkan satu variabel respon dan lebih dari satu variabel prediktor disebut regresi linier berganda.

Salah satu asumsi yang digunakan dalam suatu model regresi linier sederhana adalah bahwa setiap nilai variabel *X* berkaitan dengan suatu distribusi dari nilai-nilai variabel *Y* digunakan notasi $E(y_i|X = x_i)$ untuk menyatakan nilai harapan bersyarat bagi variabel respon Y_i untuk nilai variabel prediktor tertentu, yaitu $X = x_i$, sedangkan fungsi kepekatan variabel respon *Y* tersebut dinotasikan dengan $f(y|x)$.

Asumsi lain yang digunakan dalam suatu model regresi linier sederhana adalah bahwa hubungan antara nilai harapan bagi y_i dengan nilai x_i dapat dinyatakan melalui persamaan berikut [2]:

$$E(y_i|x_i) = \beta_0 + \beta_1 x_i \tag{1}$$

dimana β_0 dan β_1 adalah parameter regresi yang tidak diketahui nilainya. Persamaan 1 menyatakan bahwa nilai rata-rata bagi y_i untuk nilai x_i tertentu terletak dalam suatu garis lurus. Persamaan tersebut merupakan garis regresi populasi.

Model regresi dari pengamatan (y_i, x_i) dalam sampel akan memenuhi persamaan:

$$y_i = \beta_0 + \beta_1 x_i \tag{2}$$

dengan y_i merupakan nilai pengamatan dari variabel respon ke-*i* dan x_i adalah pengamatan dari variabel prediktor ke-*i*, serta β_0 dan β_i berturut-turut merupakan konstanta dan nilai koefisien regresi *X* terhadap *Y*.

Secara umum, model bagi regresi linier berganda yang melibatkan $(k-1)$ variabel *X* adalah sebagai berikut:

$$E(y_i|x_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_{k-1} x_{k-1} \tag{3}$$

dalam kasus ini, *k* adalah jumlah parameter regresi dalam model (β_i). Nilai harapan bagi variabel *Y* sama dengan β_0 jika $x_1 = x_2 = \cdots = x_{k-1} = 0$. Koefisien bagi x_k yaitu, β_k adalah perubahan dalam nilai rata-rata Y untuk setiap peningkatan x_k sebesar satu satuan jika nilai variabel *X* lainnya tetap.

Asumsi Data

Standarisasi data perlu dilakukan karena data dalam penelitian ini memiliki nilai satuan yang berbeda. Salah satu cara melakukan standarisasi data yaitu dengan menggunakan *Z-score*. *Z-score* membakukan data sehingga variabel memiliki nilai tengah nol dan ragam satu.

$$z_i = \frac{x_{ij} - \bar{x}}{sx_j\sqrt{n-1}} \text{ dimana } sx_{ij} = \sqrt{\frac{\sum_1^n (x_{ij} - \bar{x})^2}{n-1}} \quad (4)$$

Salah satu permasalahan yang perlu mendapat perhatian khusus dalam penggunaan analisis regresi linear berganda adalah kemungkinan adanya korelasi dalam variabel bebasnya. Keadaan ini biasanya terjadi ketika dalam model regresi yang digunakan terdapat suatu variabel prediktor yang berkorelasi tinggi dengan variabel prediktor lainnya. Adanya multikolinearitas mengakibatkan penduga koefisien regresi yang diperoleh dari MKT akan menghasilkan variansi yang besar, meskipun tetap tidak bias. 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:

$$VIF_i = (1 - R_i^2)^{-1} \quad (5)$$

R_i^2 merupakan koefisien determinasi dari regresi variabel prediktor x_i dengan variabel prediktor lainnya. jika nilai VIF<10 maka tidak terjadi multikolinearitas. Sebaliknya jika nilai VIF>10 maka terjadi masalah multikolinearitas.

LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR **(LASSO)**

Penduga koefisien pada metode LASSO ($\hat{\beta}^{lasso}$) diperoleh dengan cara meminimumkan persamaan berikut:

$$\hat{\beta}^{lasso} = \arg\min\left\{\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{k}\beta_j x_{ij})^2\right\} \quad (6)$$

dengan syarat $\sum_{j=1}^{k}\beta_j \le s$. Nilai s merupakan parameter *tuning* yang mengontrol penyusutan koefisien LASSO dengan $s \ge 0$. Dari Persamaan (6) dapat ditulis dalam bentuk persamaan pengali *lagrange* sebagai berikut,

$$\hat{\beta}^{lasso} = \arg\min\{\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{k}\beta_j x_{ij})^2 + \lambda\sum_{j}^{k}\left|\beta_j\right|\} \quad (7)$$

dimana y_i merupakan variabel respon pengamatan ke-*i*, β_0 dan β_j berturut-turut adalah konstanta dan koefisien regresi variabel prediktor, serta x_{ij} adalah variabel prediktor.

Nilai $s < t$ dengan $t = \sum_{j=1}^{k}\widehat{\beta}_j$ akan menyebabkan koefisien menyusut mendekati nol atau tepat nol, sehingga LASSO akan berperan sebagai seleksi variabel. Akan tetapi jika $s > t$ maka penduga koefisien LASSO memberikan hasil yang sama dengan penduga MKT [3]. Koefisien regresi LASSO ditentukan berdasarkan parameter tuning yang sudah dibakukan, yaitu $s = t \Big/ \sum_{j=1}^{k}\left|\hat{\beta}_j^0\right|$, dengan $t = \sum_{j=1}^{k}\left|\hat{\beta}_j\right|$ dimana $\hat{\beta}_j$ merupakan penduga LASSO disetiap tahapan seleksi variabel dan $\hat{\beta}_j^0$ adalah penduga MKT sedangkan nilai *s* diperoleh melalui validasi silang (CV (*cross validation*)).

Nilai λ pada Persamaan (7) disebut sebagai parameter *tuning* yang berkorespondensi satu-satu dengan *s*. Artinya untuk setiap nilai $s \ge 0$ yang menghasilkan solusi $\hat{\beta}^{lasso}$ terdapat $\lambda \ge 0$ sedemikian sehingga menghasilkan solusi $\hat{\beta}^{lasso}$ juga.

Bentuk matriks Persamaan (7) dapat dinyatakan sebagai berikut:

$$\hat{\boldsymbol{\beta}}^{lasso} = (\boldsymbol{Y} - \boldsymbol{X\beta})^T (\boldsymbol{Y} - \boldsymbol{X\beta}) + \lambda|\boldsymbol{\beta}|I \tag{8}$$

dimana,

$\boldsymbol{Y}$ = matriks pengamatan variabel dependen berukuran ($n \times 1$)
$\boldsymbol{X}$ = matriks variabel independen berukuran ($n \times k$)
$\boldsymbol{\beta}$ = matriks koefisien variabel independen berukuran ($k \times 1$)
$\boldsymbol{\lambda}$ = parameter tuning.
$|\boldsymbol{\beta}|$= matriks diagonal dengan elemen diagonal $|\beta_j|$.

Dari Persamaan (8) dapat diuraikan sebagai berikut,

$$\hat{\boldsymbol{\beta}}^{lasso} = \boldsymbol{Y}^T\boldsymbol{Y} - 2\boldsymbol{Y}^T\boldsymbol{X\beta} + \boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{X\beta} + \lambda\boldsymbol{\beta}_j \tag{9}$$

Persamaan (9) merupakan persamaan yang dapat diturunkan secara analitik jika pada matriks X berlaku sifat ortonormal. Sehingga turunan Persamaan (9) terhadap β sebagai berikut [9];

$$\hat{\boldsymbol{\beta}}^{lasso} = -2\boldsymbol{Y}^T\boldsymbol{X} + 2(\boldsymbol{X}^T\boldsymbol{X})\boldsymbol{\beta} + \lambda \boldsymbol{sign}(\boldsymbol{\beta}_j) \tag{10}$$

dengan $\boldsymbol{sign}(\boldsymbol{\beta}_j) = \begin{pmatrix} sign(\beta_{j1}) \\ \vdots \\ sign(\beta_{jk}) \end{pmatrix}$. $sign$ merupakan tanda kordinat yang memetakan entri positif ke 1, entri negatif ke -1, dan nol ke nol dari himpunan $\boldsymbol{\beta}_j$.

Dengan menetapkan Persamaan (10) sama dengan nol, diperoleh $\hat{\boldsymbol{\beta}}^{lasso}$. Asumsikan X ortonormal sehingga ($\boldsymbol{X}^T\boldsymbol{X}$) = I, maka pada kasus ortonormal taksiran MKT,

$$\hat{\boldsymbol{\beta}}_j^0 = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} = \boldsymbol{X}^T\boldsymbol{Y} \tag{11}$$

Akibatnya, taksiran $\hat{\boldsymbol{\beta}}^{lasso}$ dari Persamaan (10) diperoleh sebagai berikut,

$$\begin{aligned} \hat{\boldsymbol{\beta}}^{lasso} &= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} - \frac{\lambda}{2}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{sign}(\hat{\boldsymbol{\beta}}^{lasso}) \\ &= \hat{\boldsymbol{\beta}}_j^0 - \frac{\lambda}{2}\boldsymbol{sign}(\hat{\boldsymbol{\beta}}_j^{lasso}) \end{aligned} \tag{12}$$

Pada Persamaan (12), $\boldsymbol{sign}(\hat{\boldsymbol{\beta}}_j^{lasso})$ selalu memiliki tanda yang sama dengan $\hat{\boldsymbol{\beta}}_j^0$, artinya $\boldsymbol{sign}(\hat{\boldsymbol{\beta}}_j^{lasso}) = \boldsymbol{sign}(\hat{\boldsymbol{\beta}}_j^0)$. Sehingga Persamaan (12) menjadi,

$$\hat{\boldsymbol{\beta}}^{lasso} = (\hat{\boldsymbol{\beta}}_j^0 - \frac{\lambda}{2})\boldsymbol{I}_{[\hat{\beta}_j^0 \geq 0]} + (\hat{\boldsymbol{\beta}}_j^0 - \frac{\lambda}{2})\boldsymbol{I}_{[\hat{\beta}_j^0 \leq 0]} \tag{13}$$

sehingga taksiran penduga koefisien dengan menggunakan LASSO diperoleh,

$$\hat{\boldsymbol{\beta}}^{lasso} = \boldsymbol{sign}(\hat{\boldsymbol{\beta}}_j^0)(\hat{\boldsymbol{\beta}}_j^0 - \frac{\lambda}{2});\ j = 1, 2, 3, \ldots, k \tag{14}$$

Least Angle Regression (LARS)

Least angle regression (LARS) merupakan metode seleksi model dimana algoritmanya dapat dimodifikasi untuk diimplementasikan ke dalam penyelesaian LASSO. Modifikasi LARS menghasilkan efisiensi algoritma dalam menduga koefisien LASSO dengan komputasi yang lebih cepat dibandingkan pemrograman kuadratik.

LARS melakukan estimasi vektor prediksi $\hat{\boldsymbol{\mu}} = \boldsymbol{X}^T\hat{\boldsymbol{\beta}}$ dengan langkah-langkah yang berurutan dan di setiap langkah akan menambah satu kovarian ke dalam model [4]. Nilai vektor prediksi diperoleh dengan menggunakan nilai awal $\hat{\boldsymbol{\mu}}_0 = \mathbf{0}$, berikut adalah langkah langkah estimasi koefisien LASSO dengan LARS:

i. Mencari vektor yang sebanding dengan vektor korelasi antara variabel prediktor dengan galatnya.

$$\hat{\boldsymbol{c}} = \boldsymbol{X}^T(\boldsymbol{Y} - \hat{\boldsymbol{\mu}})$$

ii. Menentukan korelasi saat mutlak terbesar

$$\hat{C} = \max\left|\hat{c}_j\right|$$

Maka diperoleh $s_j = sign\{c_j\}$ untuk $j \in A$.

iii. Menentukan ($\boldsymbol{X}_A$), dimana himpunan A merupakan himpunan indeks aktif dari variabel prediktor. Himpunan indeks aktif A ditentukan berdasarkan nilai korelasi mutlak terbesar.

iv. Menghitung nilai vektor equingular ($\boldsymbol{u}_A$), $\boldsymbol{u}_A$ adalah suatu vektor yang membagi sudut dari kolom-kolom $\boldsymbol{X}_A$ menjadi sama besar dengan besar sudutnya kurang dari 90°. Nilai $\boldsymbol{u}_A$ diperoleh menggunakan rumus berikut:

$$\boldsymbol{u}_A = \boldsymbol{X}_A \boldsymbol{w}_A$$

dengan $\boldsymbol{w}_A$ merupakan nilai bobot dan dapat diperoleh dengan persamaan berikut:

$$\boldsymbol{w}_A = \boldsymbol{A}_A \boldsymbol{G}_A^{-1} \mathbf{1}_A\,,\ \boldsymbol{A}_A = (\mathbf{1}^T{}_A \boldsymbol{G}_A^{-1} \mathbf{1}_A)^{-\frac{1}{2}},\ \boldsymbol{G}_A = \boldsymbol{X}^T{}_A \boldsymbol{X}_A$$

v. Menghitung *inner product*

$$\boldsymbol{a} \equiv \boldsymbol{X}_A^T \boldsymbol{u}_A$$

vi. Menghitung vektor prediksi

$$\hat{\boldsymbol{\mu}}_{A+} = \hat{\boldsymbol{\mu}}_A + \hat{\gamma}\boldsymbol{\mu}_A$$

$$\hat{\gamma} = \min\nolimits^{+}_{j \in A^c}\{\frac{\hat{C} - \hat{c}}{A_A - a_A}, \frac{\hat{C} + \hat{c}}{A_A + a_A}\}$$

dimana $\hat{\gamma}$ merupakan panjang dari vektor equingular. Tanda $\min^{+}_{j \in A^c}$ menunjukkan bahwa yang dipilih adalah nilai minimum positif dari yang bukan merupakan himpunan A. Pada tahap akhir dalam memperoleh nilai $\hat{\gamma}$ menggunakan rumus $\hat{\gamma} = \frac{\hat{c}_m}{A_m}$.

Modifikasi Dari *Least Angle Regression* (LARS)

Tanda dari koordinat bukan nol $\hat{\beta}_j$ sama dengan tanda $\hat{c}_j$, dengan rumus [4]:

$$sign\left(\hat{\beta}_j\right) = sign\left(\hat{c}_j\right) = s_j = 1 \tag{15}$$

Persamaan (15) dapat diperoleh apabila koefisien bukan nol $\tilde{\gamma} = \underset{\gamma_j > 0}{min}\left\{\gamma_j\right\}$ dengan $\gamma_j = \frac{-\hat{\beta}_j}{s_j w_{Aj}}$. Apabila $\tilde{\gamma} < \hat{\gamma}$ maka $\beta_j(\gamma)$ bukan merupakan solusi LASSO karena pada $\beta_j(\gamma)$ telah berubah tanda. Sedangkan $c_j(\gamma)$ tidak berubah tanda atau dengan kata lain pembatasan tanda telah dilanggar sehingga proses LARS berhenti pada $\hat{\gamma} = 0$ dan menghapus j dari perhitungan vektor equiangular kemudian j dimasukkan kembali pada tahap perhitungan LARS selanjutnya.

VALIDASI SILANG

Validasi silang membagi data menjadi dua bagian, yaitu data *training* dan *testing*. Data *training* digunakan untuk menentukan nilai $\hat{\beta}$ atau untuk menyusun model, sedangkan data *testing* digunakan untuk menguji kebaikan $x\hat{\beta}$. Nilai validasi silang yang diperoleh merupakan penduga bagi galat prediksi. Salah satu metode tipe validasi silang adalah *k-fold*. Nilai galat prediksi *k-fold Cross validation* (CV) apabila data dipartisi menjadi c bagian diperoleh dengan persamaan berikut:

$$CV\ MSE = \frac{1}{c}\sum_{i=1}^{c}(y_i - \hat{y}_{-i(c)})^2 \qquad (16)$$

dengan $\hat{y}_{-i(c)}$ adalah dugaan y 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 5-*fold* atau 10-*fold* karena menghasilkan nilai CV dengan bias tinggi tetapi ragam rendah.

STUDI KASUS

Pada penelitian ini, variabel respon yang digunakan yaitu produk domestik regional bruto (PDRB) di wilayah Kalimantan Barat tahun 2017 [5]. Terdapat empat variabel prediktor yang diduga berpengaruh terhadap variabel respon yang diuraikan dalam tabel berikut:

Tabel 1 Variabel Respon dan Variabel Prediktor

Variabel	simbol
Produk domestik regional bruto	PDRB
Penanaman modal asing	PMA
Angkatan kerja	AK
Penanaman modal dalam negri	PMDN
Dana alokasi umum	DAU

Data dari seluruh variabel dalam penelitian ini merupakan data pada tahun 2017 di setiap kabupaten/kota wilayah Kalimantan Barat, sehingga terdapat 14 observasi (*N*). Berikut ini adalah dekripsi data dari setiap variabel pada 14 kabupaten di Kalimantan Barat pada tahun 2017.

Tabel 2 Deskripsi Statistik Data

Variabel	N	Min	Max	Mean
PDRB	14	2.185.464,90	30.480.089,3	9.417.818,94
PMA	14	89.005,00	5 .937.329,0	814.701,86
AK	14	75.799,00	467.860,00	250.634,57
PMDN	14	61.673,00	8.294.432,0	3.767.620,72
DAU	14	25.327.319,8	412.594.897,4	99.005.467,37

Dari Tabel 2 dan informasi data, diperoleh nilai PDRB tertinggi/maksimum pada Kota Pontianak sedangkan terendah/minimum pada kabupaten Kayong Utara.

Berdasarkan variabel-variabel yang mempengaruhi PDRB di wilayah Kalimantan Barat diperoleh hasil *output* model regresi menggunakan *software* R sebagai berikut:

Tabel 3 Estimasi Parameter Menggunakan MKT

Variabel	Estimasi	Std. Eror	t.hit	p-value
PMA	0,15102	0,089557	1,686	0,1227
AK	0,15437	0,083167	1,858	0,0928
PMDN	0,19687	0,065156	3,020	0,0129
DAU	0,92442	0,059771	15,463	2,61e-08

Dari Tabel 3 diperoleh hasil *p-value* pada variabel PMA dan AK $\geq 0{,}05$ sehingga dapat dikatakan bahwa variabel PMA dan AK tidak berpengaruh signifikan terhadap PDRB. Sebaliknya variabel prediktor PMDN dan DAU signifikan berpengaruh terhadap PDRB dengan nilai *p-value* $\leq$ 0,05. Berdasar estimasi parameter menggunakan MKT diperoleh model regresi sebagai berikut:

$$\hat{y} = 0{,}15102x_1 + 0{,}15437x_2 + 0{,}19687x_3 + 0{,}92442x_4$$

Uji Multikolinearitas

Uji multikolinearitas pada penelitian ini menggunakan statistik uji *variance inflation factor* (VIF). Jika nilai VIF yang diperoleh > 10, maka terjadi multikolinearitas antar variabel prediktor tersebut. Tabel 4 menunjukan *output* dari uji VIF dengan dibantu *software* R.

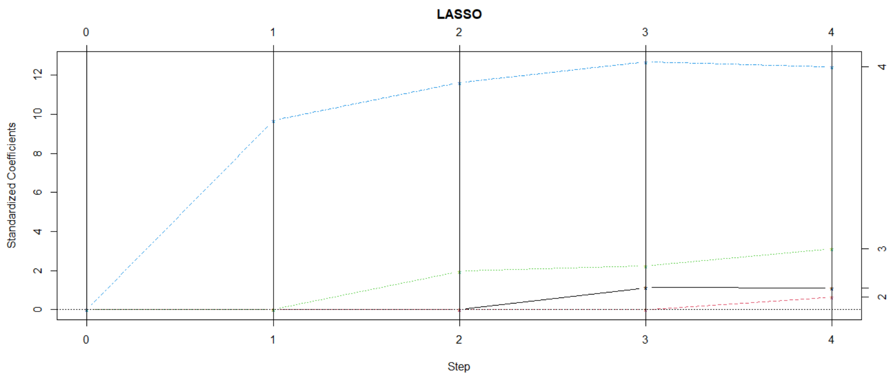
Tabel 4 Nilai VIF

Variabel	PMA	AK	PMDN	DAU
VIF	5,063	1,401	12,570	7,586

Berdasarkan Tabel 4, diketahui bahwa variabel PMDN memiliki nilai VIF >10. Sedangkan 3 variabel lainnya memiliki nilai VIF $\leq$ 10. Dapat disimpulkan terjadi masalah multikolinearitas pada model regresi tersebut, sehingga akan dilakukan penanganan dengan menggunakan metode LASSO.

Pendugaan Koefisien Menggunakan LASSO

Pendugaan koefisien LASSO dilakukan secara bertahap dengan menetapkan koefisien awal bernilai nol. Selanjutnya secara bertahap variabel prediktor yang paling berkorelasi dengan galat akan masuk ke dalam model.



Keterangan: ---- x1 ---- x2 ---- x3 ---- x4

Gambar 1 Tahapan seleksi variabel penduga LASSO dengan algoritma LARS

Pada Gambar 1 menunjukkan bahwa sumbu *Y* merupakan koefisien-koefisien metode LASSO pada setiap tahapan algoritma LARS dan sumbu *X* adalah parameter *tunning*. Dari gambar diperoleh variabel *DAU* (X_4) adalah variabel prediktor pertama yang masuk ke dalam model karena memiliki korelasi tertinggi terhadap galat. Nilai korelasi untuk variabel *DAU* adalah $C_{DAU} = 0{,}994$, kemudian penduga koefisien parameter dari DAU bergerak seiring dengan pergerakan nilai *tunning*, yakni dari $s = 0$ sampai $s = 1$.

Tabel 5 Nilai koefisien LASSO pada tahapan seleksi menggunakan LARS

PMA	AK	PMDN	DAU
0,0000000	0,0000000	0,0000000	0,0000000
0,0000000	0,0000000	0,0000000	0,7192865
0,0000000	0,0000000	0,1235974	0,8652344
0,1531799	0,0000000	0,1406429	0,9447367
0,1510115	0,1545480	0,1967430	0,9242279

Pada tahap kedua variabel AK (X_2) masuk ke dalam model ketika nilai $s = 0{,}919$. Proses ini terus berlanjut sampai semua variabel prediktor masuk ke model. Berikut ini adalah nilai koefisien dari tahapan variabel prediktor yang masuk ke metode LASSO.

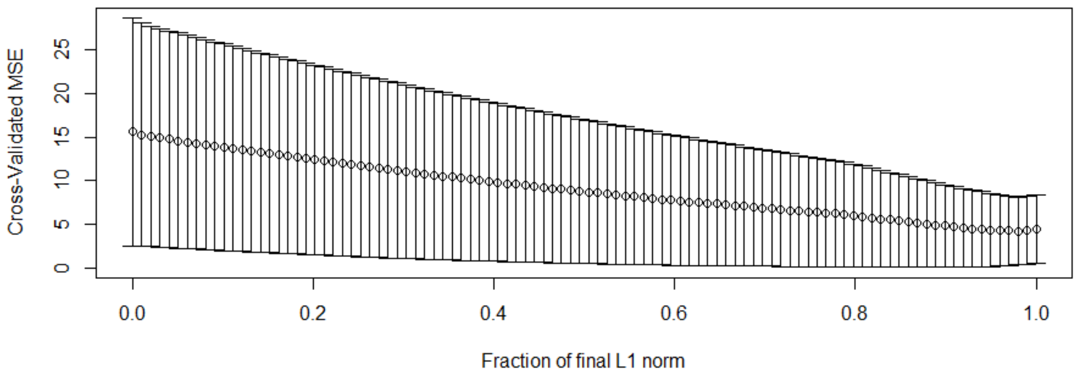
Dari Tabel 5 diperoleh koefisien awal dimulai dengan nol, selanjutnya diperoleh calon koefisien LASSO pada tahapan seleksi pertama dengan algoritma LARS yaitu:

$$\beta(\gamma)^t = \begin{pmatrix} 0.0000 & 0.0000 & 0.0000 & 0.7193 \end{pmatrix}$$

Setelah semua variabel terseleksi selanjut dilakukan pemilihan model terbaik menggunakan validasi silang.

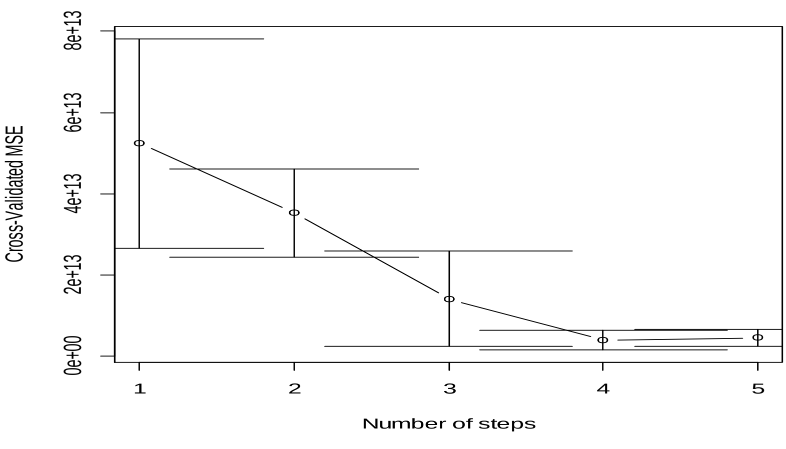
Pemilihan Model Terbaik LASSO

Pemilihan model terbaik dalam metode LASSO dilakukan dengan menggunakan kriteria validasi silang, yaitu dengan menggunakan *mode fraction* dan *mode step*. Pada *mode fraction*, nilai validasi silang dihitung berdasarkan nilai *tunning*. Pemilihan model terbaik pada *mode fraction* dilakukan dengan mengamati *plot* yang dihasilkan oleh *mode* ini.



Gambar 2 Nilai validasi silang mengunakan mode *fraction*

Pada Gambar 2 dengan sumbu Y merupakan nilai MSE dari CV dan sumbu X merupakan nilai *tunning*, diperoleh nilai *tunning* yang optimal berada pada interval $0{,}919 < s < 1$ adalah sekitar 0,919 dan merupakan nilai CV yang minimum. Nilai CV minimum tersebut dapat berbeda setiap kali melakukan pemanggilan fungsinya. Dari beberapa pengulangan maka diperoleh CV turun dan kemudian naik kembali ketika nilai $s = 1$.



Gambar 3 MSE Menggunakan CV Pada Variasi Nilai *Tuning*

Langkah selanjutnya yaitu melakukan pemilihan model terbaik LASSO dengan mode step. Perhitungan nilai validasi silang pada mode step dilakukan saat penambahan sebuah variabel prediktor disetiap tahapan. Berdasarkan hasil mode step pada Gambar 3, pemilihan model terbaik dengan menggunakan mode step dalam menghitung nilai CV menunjukkan bahwa model terbaik terlihat pada tahap ke-4. Sehingga apabila mengacu kembali pada mode *fraction* pada tabel 4.6 maka diperoleh nilai *tuning* minimum pada tahap seleki ke-4 yaitu $s = \frac{t}{\sum_{j=1}^{k}\left|\hat{\beta}_j^0\right|} = \frac{1,239}{1,426} = 0,868$, terdapat tiga variabel prediktor yang masuk ke dalam model yaitu PMA, PMDN dan DAU. Variabel prediktor AK tidak masuk ke dalam model karena memiliki penduga koefisien regresi bernilai nol dan terseleksi dari model. Berikut adalah persamaan regresi LASSO yang dihasilkan oleh perhitungan kedua mode CV:

$$\hat{y} = 0,0051 + 0,1531PMA + 0,1406\,PMDN + 0,9447DAU$$

Perbandingan Metode MKT dengan Metode LASSO

Perbandingan nilai penduga koefisien parameter dengan metode MKT dan metode LASSO dapat dilihat 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 MKT. Nilai penduga koefisien parameter untuk variabel PMDN disusutkan sampai tepat nol, sehingga variabel PMDN tidak memiliki pengaruh terhadap model.

Tabel 6 Perbandingan Nilai Parameter MKT dan LASSO

Variabel Prediktor	MKT	LASSO
PMA	0,0058	0,0058
AK	0,1510	0,1532
PMDN	0,1544	0,0000
DAU	0,9244	0,9447

Model terbaik yang dihasilkan metode LASSO dan MKT dapat ditentukan dengan nilai *Rsquare* dan MSE dari masing-masing metode. Berdasarkan Tabel 7 dapat dilihat nilai *Rsquare* metode LASSO lebih tinggi dari MKT yaitu 99,70%. Sebaliknya untuk nilai MSE metode LASSO cenderung lebih kecil dibandingkan MKT yaitu sebesar 0,2315. Berdasarkan nilai Rsquare dan MSE dari masing-masing metode dapat disimpulkan bahwa metode LASSO menghasilkan model yang lebih baik daripada MKT. Melalui penyusutan yang sampai tepat nol ini, LASSO dapat digunakan sebagai metode seleksi variabel sehingga model menjadi lebih sederhana dan efisien serta dapat mengatasi masalah multikolinieritas.

Tabel 7 Perbandingan Nilai R*square* dan MSE dari MKT dan LASSO

Model	Kriteria pemilihan model	
	Rsquare	MSE
MKT	95,54%	0,2917
LASSO	99,70%	0,2315

PENUTUP

Berdasarkan pembahasan yang telah dilakukan, dapat disimpulkan bahwa model regresi terbaik menggunakan metode LASSO diperoleh pada tahap keempat pada saat nilai $s = 0,868$, yaitu dengan model sebagai berikut:

$$\hat{y} = 0{,}1531PMA + 0{,}1406PMDN + 0{,}9447DAU$$

Multikolinearitas yang terjadi pada variabel PMDN (X_3) telah teratasi menggunakan metode LASSO.

DAFTAR PUSTAKA

[1]. Pradewi, E. D., & Sudarno. (2012). Kajian Estimasi-M IRLS Menggunakan Fungsi Pembobot Huber dan Bisquare Tukey Pada Data Ketahanan Pangan di Jawa Tengah. *Media Statistika*, **5** (1): 1–10.

[2]. Debataraja, Naomi N., Kusnandar, D., Mara, Muhlasah N., Satyahadewi, Neva. (2019)*Metode Statistik Serta Aplikasinya dengan MINITAB, Excel dan R*, UNTAN Press, Pontianak.

[3]. Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of The Royal Statistical Society Series B Methodological*, **58** (1): 267-288.

[4]. Efron, B., Hastie, T., Johnstone, I., Tibshirani, R. (2004). Least Angle Regression. *The Annals of Statistics*, **32** (2): 407-499

[5]. Badan Pusat Statistik. 2011. Pertumbuhan Ekonomi Kalimantan Barat 2018. *http://www.bps.go.id/*.

[6]. Andana, Aulia Putri, 2017. Model Regresi Menggunakan Least Absolute Shringkage and Selection Operator Pada Data Banyaknya Gizi Buruk Kabipaten/Kota di Jawa Tengah. *Jurnal Gaussian,* **6** (1): 21-31

[7]. Prabowo, F. K., Rusgiyono, A., & Wilandari, Y. (2015). Pemodelan Pertumbuhan Ekonomi Jawa Tengah Menggunakan Pendekatan Least Absolute Shrinkage and Selection Operator (LASSO). *Jurnal Gaussian,* **4** (4): 855-864.

[8]. Pusporini, Arum. (2012). Penerapan Regresi Gulud dan Least Absolute Shrinkage and Selection Operator (LASSO) dalam Penyusutan Koefisien Regresi. *Bogor: Institut Pertanian Bogor*.

[9]. Robbani, Muhammad. (2019). Regresi Least Absolute Shrinkage and Selection Operator (LASSO) Pada Kasus Inflasi Di Indonesia Tahun 2014-2017. *Jurnal EurekaMatika,* **7** (2).

OKTAVIANUS FRANS : Jurusan Matematika FMIPA Untan, Pontianak
h1011141031@student.untan.ac.id

SETYO WIRA RIZKI : Jurusan Matematika FMIPA Untan, Pontianak
setyo.wirarizki@math.untan.ac.id

DADAN KUSNANDAR : Jurusan Matematika FMIPA Untan, Pontianak
dkusnand@untan.ac.id