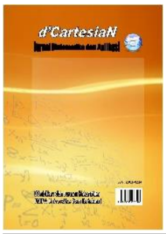


d'CartesiaN

Jurnal Matematika dan Aplikasi

p-ISSN:2302-4224 e-ISSN: 2685-1083

Journal homepage: https://ejournal.unsrat.ac.id/index.php/decartesian

Kajian Model Prediksi Metode *Least Absolute Shrinkage and Selection Operator* (LASSO) pada Data Mengandung Multikolinearitas

Yeremia Armstrong Mait[1], Deiby Tineke Salaki[1*], Hanny A. H. Komalig[1]

[1]Jurusan Matematika–Fakultas Matematika dan Ilmu Pengetahuan Alam–Universitas Sam Ratulangi Manado, Indonesia

[*]Corressponding author : deibyts.mat@unsrat.ac.id

A B S T R A K

Indonesia dan semua negara di dunia masih berkutat dengan penanganan masalah Covid-19. Pengeluaran anggaran pemerintah terhadap penanganan di bidang kesehatan serta penyaluran bantuan sosial untuk masyarakat dapat mempengaruhi tingkat inflasi di Indonesia. Analisis yang digunakan adalah analisis regresi. Metode kuadrat terkecil (MKT) merupakan salah satu metode pendugaan parameter yang mensyaratkan ketiadaan multikolinearitas antar peubah bebas. Jika terdapat multikolinearitas maka *Least Absolute Shrinkage and Selection Operator* (LASSO) merupakan metode alternatif yang dapat menyeleksi peubah sekaligus menduga parameter. Tujuan penelitian ini adalah membandingkan performa LASSO dengan MKT dilihat dari *root mean squared error* (RMSEP) dan nilai *mean absolute error* (MAE) serta menentukan model regresi inflasi sejumlah kota besar di Indonesia dengan menggunakan metode LASSO. Data yang digunakan adalah data simulasi yang dibangkitkan dengan bantuan aplikasi dan data real yaitu inflasi 90 kota besar di Indonesia tahun 2020. Hasil penelitian menunjukkan bahwa LASSO menghasilkan nilai RMSEP dan nilai MAE yang jauh lebih baik dari MKT serta LASSO dapat menghasilkan model regresi untuk inflasi sejumlah kota besar di Indonesia dengan baik.

INFO ARTIKEL

Diterima :

Diterima setelah revisi :

Tersedia *online* :

Kata Kunci:
Inflasi
LASSO
Metode kuadrat terkecil
Multikolinearitas

A B S T R A C T

Indonesia and all countries in the world are still struggling with the handling of the Covid-19 problem. The government's budget expenditure on handling in the health sector and the distribution of social assistance to the community can affect the inflation rate in Indonesia. The analysis used is regression analysis. The least squares method is a parameter estimation method that requires the absence of multicollinearity between independent variables. If there is multicollinearity, the Least Absolute Shrinkage and Selection Operator (LASSO) is an alternative method that can select variables as well as estimate parameters. The purpose of this study was to compare the performance of LASSO with MKT in terms of the root mean squared error (RMSEP) and the mean absolute error (MAE) and to determine the inflation regression model for a number of major cities in Indonesia using the LASSO method. The data used are simulation data generated with the help of applications and real data, namely inflation in 90 major cities in Indonesia in 2020. The results show that LASSO produces RMSEP and MAE values which are much better than least square method and LASSO can produce regression models for inflation of a number of big cities in Indonesia.

ARTICLE INFO

Accepted :

Accepted after revision :

Available online :

Keywords:
Inflation
LASSO
Ordinary least square
Multicolinearity

1. PENDAHULUAN

Indonesia dan semua negara di dunia masih berkutat dengan penanganan masalah Covid-19 khususnya dalam masalah perekonomian.

Seiring dengan terjadinya pandemi ini menyebabkan realisasi belanja pemerintah pusat diprioritaskan pada bidang kesehatan, bantuan sosial, dan pemulihan ekonomi negara. Pengeluaran anggaran pemerintah terhadap penanganan di bidang kesehatan serta penyaluran bantuan sosial untuk masyarakat dapat mempengaruhi tingkat inflasi di Indonesia [1].

Inflasi adalah peningkatan harga berbagai jenis barang secara terus menerus serta menurunkan nilai uang yang cenderung terjadi pada negara berkembang seperti Indonesia. Salah satu faktor yang sangat berpengaruh pada tinggi rendahnya inflasi adalah indeks harga konsumen (IHK).

Indeks Harga Konsumen (IHK) adalah salah satu indikator penting dalam perekonomian yang dapat memberikan informasi terkait perkembangan harga barang dan jasa yang dibayarkan oleh konsumen dalam hal ini masyarakat [2].

Metode kuadrat terkecil (MKT) adalah metode yang sering digunakan untuk menduga parameter model tersebut. Namun demikian, MKT memiliki asumsi-asumsi ketat yang harus dipenuhi. Salah satunya adalah ketiadaan multikolinearitas atau korelasi tinggi antar peubah bebas [3].

Least Absolute Shrinkage and Selection Operator (LASSO) merupakan salah satu metode alternatif yang dapat menangani multikolinearitas. Solusi LASSO diperoleh dengan meminimumkan jumlah kuadrat galat terhadap suatu kendala yang selain menduga parameter juga dapat menyusutkan koefisien regresi menjadi tepat nol sehingga dapat digunakan sebagai seleksi peubah. pada jumlah data yang berukuran relatif besar, LASSO menghasilkan model yang lebih baik dari MKT [4].

Penelitian ini mengkaji performa model regresi LASSO dibandingkan dengan MKT. Data yang digunakan adalah data simulasi dan juga data real untuk menentukan faktor-faktor yang mempengaruhi terjadinya inflasi di 90 kota besar di Indonesia pada tahun 2020.

Analisis Regresi Linear

Analisis regresi linear merupakan suatu teknik untuk memodelkan hubungan linear antar peubah. selain dapat mengetahui hubungan antara peubah bebas dan peubah respon, persamaan regresi dapat juga digunakan untuk meramalkan nilai-nilai suatu peubah respon dari nilai-nilai satu atau lebih peubah bebas. Jika terdapat sejumlah p peubah bebas, $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_p$ dan peubah respon $\boldsymbol{y}$, maka model regresi linier memiliki bentuk sebagai berikut [5]

$$y_i = \beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j + \varepsilon_i \qquad (1)$$

Metode Kuadrat Terkecil

Metode kuadrat terkecil (*least squares method*) merupakan salah satu metode yang digunakan untuk menghitung koefisien regresi sampel ($\widehat{\boldsymbol{\beta}}$) sebagai penduga bagi koefisien regresi populasi ($\boldsymbol{\beta}$) dengan meminimumkan jumlah kuadrat galat (JKG). Berdasarkan persamaan (1), galat $\boldsymbol{\varepsilon}$ dapat dituliskan sebagai $\boldsymbol{\varepsilon} = \boldsymbol{y} - \boldsymbol{X\beta}$ dan JKG adalah

$$\begin{aligned} &\varepsilon'\varepsilon = (\mathrm{y} - \mathrm{X}\beta)'(\mathrm{y} - \mathrm{X}\beta) \qquad \Big| \ \frac{\partial(\varepsilon'\varepsilon)}{\partial\beta} \\ &\mathrm{X}'\mathrm{X}\beta = \mathrm{X}'\mathrm{y} \end{aligned} \qquad (2)$$

Jika $\boldsymbol{X'X}$ merupakan matriks non singular, maka penduga kuadrat terkecil untuk $\boldsymbol{\beta}$ adalah :

$$\hat{\beta}^{\mathrm{MKT}} = (\mathrm{X}'\mathrm{X})^{-1}(\mathrm{X}'\mathrm{y}) \qquad (3)$$

Pada kondisi terjadi multikolinearitas antar vektor kolom penyusun matriks $\boldsymbol{X}$ maka matriks $\boldsymbol{X'X}$ akan menghasilkan akar karakteristik yang relatif kecil bahkan hampir nol. Adanya multikolinearitas juga dapat menyebabkan det $(\boldsymbol{X'X}) \approx 0$ sehingga matriks $\boldsymbol{X'X}$ cenderung singular. Hal ini mengakibatkan nilai $\widehat{\boldsymbol{\beta}}$ yang dihasilkan oleh MKT menjadi tidak stabil [3].

Multikolinearitas

Multikolinearitas adalah situasi adanya korelasi atau hubungan kuat antar peubah bebas dalam regresi linear berganda. Penentuan adanya multikolinearitas antara peubah $\boldsymbol{x}_i$ dan $\boldsymbol{x}_j$ dilakukan dengan pemeriksaan terhadap koefisien korelasi r_{ij}. Jika $\boldsymbol{x}_i$ dan $\boldsymbol{x}_j$ mempunyai hubungan linear yang erat, maka $|r_{ij}|$ yang mengindikasikan korelasi berpasangan dari peubah $\boldsymbol{x}_i$ dan $\boldsymbol{x}_j$ akan mendekati satu.

Multikolinearitas dapat juga dideteksi menggunakan nilai Variance Inflation Factor (VIF). Nilai VIF dapat dicari menggunakan Persamaan berikut.

$$VIF_{(j)} = \frac{1}{(1 - R_j^2)} \quad j = 1,2,\dots,p \qquad (4)$$

dimana R_j^2 merupakan koefisien determinasi dari peubah X_j. Jika nilai $VIF(j)$ lebih besar dari 5 maka terjadi masalah multikolinearitas [6].

LASSO (*Least Absolute Shrinkage and Selection Operator*)

Penduga parameter regresi dengan metode ini diperoleh dengan meminimumkan jumlah kuadrat galat dengan penambahan kendala tertentu. Kendala ini memungkinkan LASSO melakukan seleksi peubah dengan cara menyusutkan koefisien regresi dari peubah bebas yang memiliki korelasi tinggi dengan galat menjadi tepat pada nol atau mendekati nol [4].

LASSO merupakan salah satu metode yang termasuk dalam regresi berkendala (*penalized regression*). Jika solusi MKT diperoleh dengan meminimumkan JKG pada persamaan (2.5), maka regresi berkendala akan menambahkan komponen λ sebagai berikut:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X'X} + \boldsymbol{\lambda I})^{-1}(\boldsymbol{X'y}) \qquad (5)$$

Dalam bentuk formulasi minimisasi berkendala, persamaan (2.4) dapat dituliskan dalam bentuk

$$\hat{\beta}^{LASSO} = \underset{\beta}{min}(\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta}) + \lambda \sum_{1}^{p} |\beta_j| \qquad (6)$$

Kendala $\lambda \sum_{j=1}^{p} |\beta_j|$ pada LASSO dalam persamaan (6) dikenal sebagai kendala L_1 karena $\sum |\beta_j| = \|\boldsymbol{\beta}\|_1$, yaitu jumlah absolut dari parameter $\boldsymbol{\beta}$. Penggunaan kendala ini menghasilkan solusi dimana sejumlah parameternya bernilai nol.

Parameter $\lambda \geq 0$ disebut *tunning parameter* yang mengontrol besarnya penyusutan nilai parameter regresi. Besaran nilai λ ini menyebabkan peubah yang berpengaruh terhadap model akan terpilih dan sebaliknya yang berpengaruh kecil akan tereliminasi. Semakin besar nilai λ akan semakin banyak parameter yang bernilai nol. Nilai λ ini juga menjadi perimbangan antara ragam dan bias. Bias akan semakin besar dengan bertambahnya nilai λ dan sebaliknya ragam akan semakin besar seiring berkurangnya nilai λ [4].

LAR (*Least Angle Regression*)

Solusi parameter LASSO pada persamaan (2.4) dapat diselesaikan dengan pemrograman kuadratik. *Least Angle Regression* (LAR) merupakan algoritma yang lebih efisien yang dimodifikasi untuk menduga parameter tersebut [7]. Algoritma LAR diuraikan sebagai berikut:

1. Membakukan peubah bebas sehingga memiliki rataan 0 dan ragam 1.
2. Menetapkan penduga koefisien regresi $\beta_1 = 0, \beta_2 = 0, \dots, \beta_p = 0$.
3. Memasukkan peubah bebas yang berkorelasi tertinggi dengan sisaan ke dalam model. Jika koefisien yang bukan nol mencapai nol, keluarkan peubah dari kumpulan peubah-peubah aktif dan arah kuadrat terkecil dihitung kembali.
4. Lanjutkan hingga *p* peubah bebas dimasukkan.
5. Gunakan validasi silang untuk memperoleh solusi model akhir LASSO.

Validasi Silang (*Cross Validation*)

Validasi silang merupakan metode untuk mengestimasi galat prediksi dalam meningkatkan ketepatan dari pemilihan model. Validasi silang atau *cross-validation* (CV) bekerja dengan membagi data menjadi dua bagian, yaitu data training dan data test. Validasi silang digunakan untuk memprediksi model dan memperkirakan seberapa akurat sebuah model. Tujuan validasi silang adalah untuk mendefinisikan dataset untuk menguji model pada tahap training supaya dapat membatasi terjadinya overfitting [8].

Validasi silang berjalan dengan data dibagi dalam 2 subset yang dinamakan data *training* dan data *testing*. Pada setiap lipatan (*fold*) dilakukan analisis atau training pada $k-1$ bagian data yaitu data *training* dan validasi dilakukan pada 1 bagian lainnya yaitu data *testing*. Proses validasi silang diulang sampai k kali. Validasi silang menghasilkan ukuran *root mean squared error prediction* (RMSEP) untuk kinerja model prediksi [8].

Kriteria Kebaikan Model

Suatu model statistik dapat dikatakan sebagai model yang baik apabila memenuhi beberapa kriteria berikut :

1. Mempunyai identifikasi tinggi pada data yang dianalisis sehingga parameter-parameter $\beta_1, \beta_2, \beta_3, \dots, \beta_n$ yang diestimasi akan memiliki nilai yang berdiri sendiri.
2. Model regresi semakin baik apabila memiliki nilai galat (*error*) yang sangat kecil. Meski data yang diamati tidak dapat dimodelkan secara sempurna, namun dapat didekati tingkat kebaikan model dengan meminimumkan jumlah galat [9].

2. METODE PENELITIAN

Waktu dan Tempat Penelitian

Penelitian dilaksanakan selama bulan Januari-Juni tahun 2020 dengan tempat penelitian di rumah atau *work from home* karena adanya pandemi Covid-19.

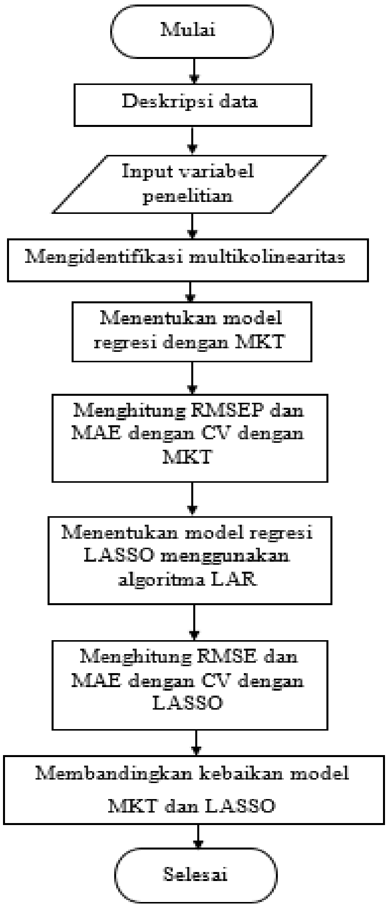
Sumber data

Dalam penelitian ini dilakukan analisis pada data bangkitan atau data simulasi dan data real. Data simulasi dibangkitkan sebanyak $n = 30$, peubah bebas $X_1, X_2, \dots, X_{100}$, $\boldsymbol{\beta} = \beta_1, \beta_2, \dots, \beta_{100}$ serta peubah respon $y_1, y_2, \dots, y_{30}$. Sedangkan data real adalah data sekunder diambil di website resmi BPS Indonesia, data tingkat inflasi di sejumlah kota besar di Indonesia 2020 dengan $n = 90$, peubah bebas $X_1, X_2, \dots, X_{20}$ serta $y_1, y_2, \dots, y_{90}$ [10].

Tahapan Penelitian

1. Mendeskripsikan data
2. Mengidentifikasi multikolinearitas antar peubah bebas
3. Menentukan model regresi linear berganda dengan menggunakan Metode kuadrat terkecil (MKT).
4. Mengevaluasi kebaikan model MKT dengan menghitung nilai *Root mean square error prediction* (RMSEP) dan *Mean absolute error* (MAE) menggunakan *cross validation* (CV).
5. Menentukan model regresi linear berganda menggunakan regresi *Least absolute shrinkage and selection operator* (LASSO) dengan algoritma Least Angle Regression (LAR).
6. Melakukan *cross validation* (CV) untuk mengevaluasi kebaikan model *Least absolute shrinkage and selection operator* (LASSO) dengan menghitung nilai *Root mean square error prediction* (RMSEP) dan *Mean absolute error* (MAE).
7. Membandingkan kebaikan model MKT dan LASSO.

Diagram Alir (*Flowchart*)



Gambar 1. Diagram Alur Penelitian

3. HASIL DAN PEMBAHASAN

Multikolinearitas pada Data

Data simulasi diatur dengan nilai korelasi sebesar $Cor = 0{,}95$ yang mengartikan bahwa data yang dibangkitkan mengandung multikolinearitas. Data real inflasi sejumlah kota besar di indonesia diuji dengan menggunakan *Variance Inflation Factor* (VIF). Hasil uji multikoliearitas data real ditunjukkan pada Tabel 1.

Tabel 1. Uji Multikolinearitas Data Real

Peubah bebas	Nilai VIF
x_1	1,355
x_2	27,818
x_3	19,039
x_4	3,398
x_5	6,073
x_6	5,157
x_7	3,648
x_8	21,141
x_9	7,426
x_{10}	9,964
x_{11}	16,704
x_{12}	6,896
x_{13}	4,946
x_{14}	4,001
x_{15}	27,298
x_{16}	3,324
x_{17}	4,423
x_{18}	6,254
x_{19}	2,972
x_{20}	4,475

Model Regresi Kuadrat Terkecil

Pendugaan model regresi dengan MKT dilakukan dengan CV. Data dipartisi menjadi dua bagian, data *training* yang mencakup 80% dari data keseluruhan dan sisanya sebagai 20% sebagai data *testing*.

Penggunaan MKT pada data mengandung multikolinearitas memberikan hasil yang tidak konsisten dimana meski nilai R^2 cukup tinggi namun tidak ada peubah bebas yang secara signifikan berpengaruh terhadap peubah respon. Dugaan koefisien dengan MKT dapat dilihat pada Persamaan (7) untuk data simulasi dan Persamaan (8) untuk data real.

$$\hat{Y} = 3{,}353x_1 - 0{,}491x_2 - 1{,}246x_3 - 76{,}719x_4 + \cdots + 0{,}389x_{78} \tag{7}$$

$$\hat{Y} = 4{,}586x_1 + 1{,}331x_2 + 0{,}421x_3 - 3{,}969x_4 + \cdots + 10{,}418x_{20} \tag{8}$$

Kebaikan model diukur berdasarkan nilai *root mean square error prediction* (RMSEP) dan *mean absolute error* (MAE). Metode kuadrat terkecil pada data simulasi menghasilkan nilai RMSEP dan MAE yang besar sesuai Tabel 2. Nilai ini relatif sangat besar yang mengindikasikan besarnya galat atau buruknya hasil pengepasan (*fitting*) dari model yang dihasilkan MKT.

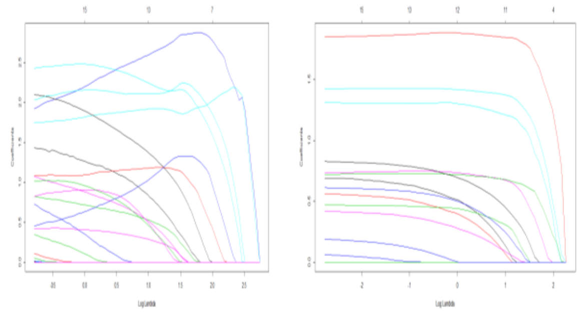
Tabel 2. Nilai galat dengan MKT

	RMSEP	MAE
Data Simulasi	654.586,2	652.316,9
Data Real	28,298	19,862

Model Regresi LASSO

Metode LASSO menggunakan algoritma LAR sesuai langkah-langkah pada LAR. Semua peubah bebas dibakukan sehingga berdistribusi $N(0,1)$. Selanjutnya menetapkan nilai-nilai koefisien $\boldsymbol{\beta}$ mulai dari nol. Data amatan baik data simulasi maupun data real dipartisi menjadi dua bagian yaitu 80% dari data keseluruhan menjadi *data training* dan 20% sebagai *data testing*.

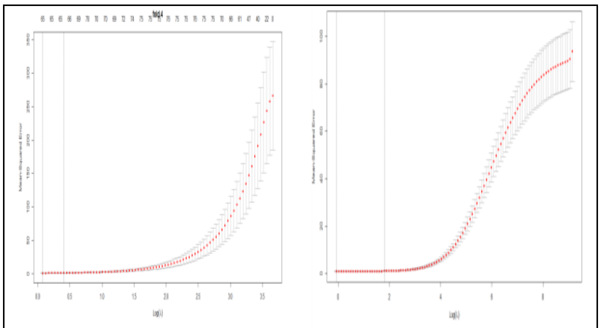
Peubah bebas yang berkorelasi tinggi dimasukkan ke dalam model sampai semua peubah bebas masuk dalam model. Koefisien regresi dapat diperoleh menggunakan nilai tuning parameter yaitu λ optimal.



(a) Data Simulasi (b) Data Real

Gambar 2. Proses Penyusutan koefisien

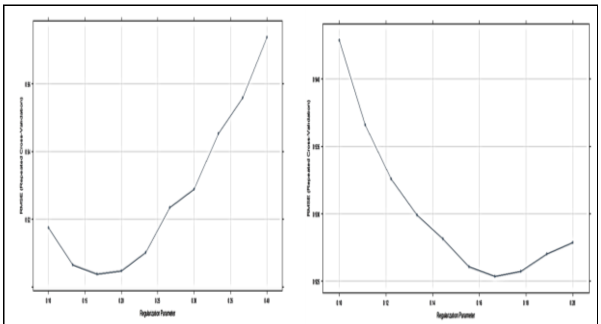
Gambar 2 menjukkan proses penyusutan koefisien regresi yang ditetapkan mulai dari nol. Penentuan nilai tuning parameter ditentukan dengan CV *10-folds*. Setiap langkah LAR penentuan tuning parameter, diperoleh langkah LAR paling optimal yaitu pada *fold-4* untuk data simulasi karena menghasilkan nilai galat paling rendah yaitu 2,751. Sedangkan data real diambil *fold-1* karena memiliki nilai galat terkecil yaitu 1,225. Penentuan tuning parameter ditunjukkan pada Gambar 3.



(a) Data Simulasi (b) Data Real

Gambar 3. Penentuan Tuning Parameter (λ)

Berdasarkan Gambar 3 dapat dilihat bahwa nilai tuning parameter paling optimal dapat diperoleh pada selang nilai $0 < \lambda < 0,5$ untuk data simulasi dan $0 < \lambda < 2$ untuk data real masing-masing sesuai Gambar 3 (a) dan Gambar 3 (b). Proses penentuan nilai λ optimal dicari dengan menetapkan 10 titik (*length*) lambda.



(a) Data Simulasi (b) Data Real

Gambar 4. Plot Penentuan λ

Gambar 4 adalah perbesaran dari Gambar 2 untuk menetapkan nilai λ optimal untuk menduga koefisien regresi dan menghasilkan galat optimal. Titik terendah pada Gambar 4 (a) adalah $\lambda = 0,168$ sedangkan pada Gambar 4 (b) adalah $\lambda = 0,16$. Artinya pada nilai lambda tersebut LASSO akan menghasilkan nilai dugaan dan nilai galat paling optimal, sehingga diambil nilai tuning parameter 0,168 untuk data simulasi dan 0,16 untuk data real karena nilai lambda tersebut menghasilkan nilai galat terkecil sesuai Gambar 4.

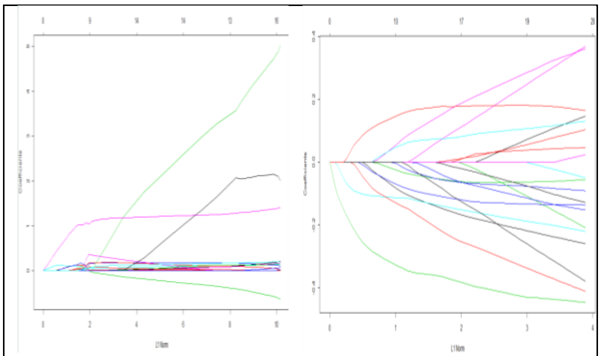
Nilai dugaan koefisien terbaik yang dihasilkan LASSO ditunjukkan pada Persamaan (9) dan Persamaan (10).

$$\hat{Y} = 0,0x_1 + 0,133x_{17} + 0,114x_{32} + 0,0x_{56} + \cdots + 0,005x_{99} \tag{9}$$

$$\hat{Y} = 0,133x_1 + 0,0x_2 - 0,044x_3 - 0,026x_4 + \cdots + 0,157x_{20} \tag{10}$$

Persamaan (9) adalah hasil dugaan regresi LASSO pada data simulasi dan Persamaan (10) adalah hasil dugaan LASSO pada data inflasi atau data real. LASSO bekerja dengan algoritma LAR untuk menyusutkan koefisien peubah bebas mendekati nol bahkan tepat nol. Hal ini dapat dilihat pada Persamaan (9) dan (10) dimana nilai koefisien sejumlah peubah bebas menyusut tepat nol dan koefisien yang lain mendekati nol.

Penyusutan setiap koefisien untuk model terbaik LASSO pada Persamaan (9) dan (10) ditampilkan dengan plot visual model final LASSO pada Gambar 5 untuk data simulasi dan data real.



(a) Data Simulasi (b) Data Real

Gambar 5. Model Final Regresi LASSO

Hal ini menunjukkan bahwa, LASSO dapat melakukan seleksi peubah. Peubah bebas yang tidak berpengaruh terhadap peubah respon, parameter dugaannya akan menyusut ke nol. Algortima LAR yang digunakan melakukan input peubah pada setiap peubah bebas sehingga dengan nilai tuning parameter optimal menghasilkan nilai dugaan koefisien regresi paling optimal. Nilai kebaikan model yang dihasilkan LASSO dapat dilihat pada Tabel 3.

Tabel 3. Nilai Galat dengan LASSO

	RMSEP	MAE
Data Simulasi	0,503	0,44
Data Real	0,925	0,747

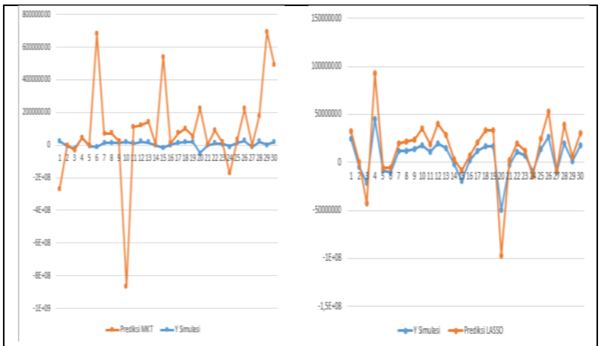
Nilai galat yang dihasilkan LASSO pada Tabel 3 menunjukkan bahwa LASSO dapat menghasilkan model regresi yang baik. Hal ini dapat dilihat pada nilai RMSEP dan MAE yang sangat optimal mendekati 0 pada data simulasi maupun data real.

Perbandingan Model MKT dan LASSO

Tabel 4. Kebaikan model MKT dan LASSO

		RMSEP	**MAE**
Data Simulasi	MKT	654586,2	652317
	LASSO	0,503	0,44
Data Real	MKT	28,298	19,862
	LASSO	0,925	0,747

Berdasarkan Tabel 4, kemampuan model LASSO untuk memprediksi lebih baik dari pada MKT. Hal ini ditunjukkan oleh nilai RMSEP dan MAE yang dihasilkan LASSO jauh lebih kecil dibandingkan dengan MKT. Artinya, LASSO jauh lebih baik memodelkan model regresi dibandingkan dengan MKT baik pada data simulasi maupun pada data inflasi sejumlah kota di Indonesia tahun 2020. Perbandingan nilai prediksi dengan metode LASSO dan MKT dan nilai aktual dinyatakan pada Gambar 6 untuk MKT dan Gambar 7 untuk LASSO.

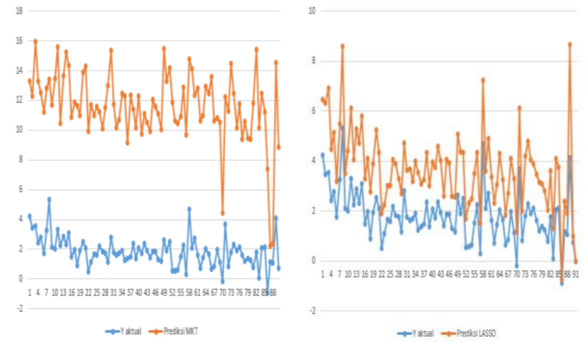


(a) MKT (b) LASSO

Gambar 6. Daya Ramal MKT dan LASSO pada Data Simulasi

Berdasarkan Gambar 6 (a), dapat dilihat bahwa plot prediksi MKT dan data simulasi memiliki *trend* yang cenderung berbeda. Hal ini sesuai dengan nilai korelasi antara keduanya yang hanya sebesar 0.182. Hasil prediksi LASSO sangat berbeda dengan hasil prediksi MKT. Jika dilihat dari Gambar 6 (b) *trend* nilai prediksi LASSO dan data simulasi relatif sama serta memiliki pola sebaran yang bersesuaian sebagaimana dengan nilai korelasi antar keduanya sebesar 0,889.

Sebaran hasil prediksi yang berwarna oranye relatif sama dengan data simulasi yang berwarna biru pada Gambar 6 (b) sedangkan Sebaran hasil prediksi yang berwarna oranye berbeda dengan data simulasi yang berwarna biru sebagaimana pada Gambar 6 (a).



(a) MKT (b) LASSO

Gambar 7. Daya Ramal MKT dan LASSO pada Data Real (inflasi)

Plot nilai aktual dan nilai prediksi menggunakan MKT pada Gambar 7 (a) menyebar dengan tren yang cenderung berbeda dan berjauhan dengan koefisien korelasi hanya sebesar $-0{,}074$. Hal ini menunjukkan tren keduanya relatif tidak sama. Sedangkan Gambar 7 (b) menunjukkan menunjukan tren yang bersesuaian antara nilai prediksi LASSO dengan nilai data real. Nilai prediksi LASSO dan nilai real memiliki koefisien korelasi sebesar 0,772. Hal ini menunjukkan keduanya memiliki tren yang relatif sama.

Perbandingan daya ramal antara MKT dan LASSO yang dinyatakan pada Gambar 6 dan Gambar 7 menunjukkan kondisi dimana LASSO lebih baik dalam memprediksi nilai y aktual pada data simulasi maupun data inflasi atau data real bila dibandingkan dengan MKT. Diperkuat dengan nilai koefisien korelasi model LASSO yang mendekati 1 sementara dengan MKT memiliki nilai korelasi relatif sangat kecil.

4. PENUTUP

Kesimpulan

Berdasarkan hasil penelitian yang sudah dilakukan, diperoleh kesimpulan sebagai berikut.

1. Berdasarkan nilai *root mean square error* (RMSEP) dan *mean absolute error* (MAE) model prediksi LASSO memiliki performa yang lebih baik dari MKT.
2. Model regresi untuk inflasi 90 kota besar di Indonesia adalah sebagai berikut :

$$\hat{Y} = 0{,}113x_1 - 0{,}044x_3 - 0{,}026x_4 - 0{,}112x_5 + 0{,}093x_6 - 0{,}096x_{10} + 0{,}003x_{12} + 0{,}136x_{14} - 0{,}318x_{15} + 0{,}040x_{17} - 0{,}008x_{19} + 0{,}157x_{20} \quad (11)$$

Saran

Penelitian selanjutnya pada data dengan multikolinearitas dapat dilakukan dengan menggunakan model regresi berkendala lainnya seperti Ridge dan *partial least squared*.

REFERENSI

[1] Kementrian Keuangan RI. 2020. *Pemerintah Waspada Dampak Pandemi Covid-19 Terhadap Ekonomi Indonesia*. SP-27/KLI/2020.

[2] Robbani M. 2019. *Regresi Least Absolute Shrinkage and Selection Operator (LASSO) pada Kasus Inflasi di Indonesia Tahun 2014-2017.* Jurnal EurekaMatika.Vol. 7(2):No 1-16.

[3] Montgomery, D.C & Peck. 1992. *Introduction to Linear Regression Analysis*. New York. John Wiley & sons.

[4] Tibshirani R. 1996. *Regression Shrinkage and selection via the LASSO*. Journal of the Royal Statistical Society, Series B. Wiley. 58(1): 267-88.

[5] Draper, N.R dan Smith, H. 1992. *Analisis Regresi Terapan*. Ed ke-2. Terjemahan oleh Bambang Sumantri. Gramedia Pustaka Utama, Jakarta.

[6] Gujarati, D. N. and Porter, D. C. 2009. *Basics Econometrics*. Ed. ke-5. McGraw-Hill/Irwin, New York.

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

[8] Salaki, D. T. 2018. *A Simulation Study of Model Averaging In High Dimensional Data With Various Correlation Structures Using Ridge Regression Methods*. Disertasi. Program Pasca Sarjana Institut Pertanian Bogor. Bogor.

[9] Gujarati, D. 2006. *Ekonometrika Dasar*. Diterjemahkan oleh Sumarto Zain. Erlangga. Jakarta.

[10] Badan Pusat Statistik. 2021. *Indikator Ekonomi Februari 2021*. Statistics Indonesia. Jakarta.

Yeremia Armstrong Mait (17101103024@student.unsrat.ac.id)

Lahir di Tomohon, 30 Juli 1999. Menempuh pendidikan tinggi Jurusan Matematika, FMIPA, Universitas Sam Ratulangi Manado. Tahun 2021 adalah tahun terakhir ia menempuh studi. Makalah ini merupakan hasil penelitian skripsinya yang dipublikasikan.

Deiby Tineke Salaki (deibyts.mat@unsrat.ac.id)

Lahir di Minahasa Selatan, 17 Desember 1972. Gelar Sarjana Matematika diperoleh tahun 1998 di Jurusan Matematika IPB Bogor. Tahun 2009 menyelesaikan studi S2 di Jurusan Matematika IPB. Tahun 2018 menyelesaikan studi S3 pada bidang Matematika di IPB. Saat ini menjadi pengajar tetap di Jurusan Matematika FMIPA UNSRAT Manado.

Hanny A. H. komalig (hannkomal@gmail.com)

Lahir pada 6 Maret 1968. Pada tahun 1991 memperoleh gelar sarjana di Universitas Sam Ratulangi Manado. Gelar Magister Sains diperoleh dari Institut Pertanian Bogor pada tahun 1999. Ia bekerja di Unsrat di Program Studi Matematika sebagai pengajar akademik tetap di Jurusan Matematika FMIPA UNSRAT Manado.