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

(Skripsi)

Oleh

Putri Maulidina Fadilah

JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
2018

ABSTRACT

REGRESSION ANALYSIS USING *LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR* (LASSO) METHOD TO HANDLE MULTICOLINEARITY

By

PUTRI MAULIDINA FADILAH

LASSO can solve multicolinearity by shrinkage some coefficients to zero or exactly zero. With LARS algorithm the calculation of LASSO is more efficient than quadratic programming. The purpose of this study is to know the performance of LASSO to handle multicolinearity and compare its etimates with OLS. The results show that LASSO can not solve the multicolinearity when all of the independent variables are correlated. However, LASSO solve the multicolinearity when the correlation partially exist in the independent variables. In higher sample sizes, LASSO gives regression coefficients estimation better than OLS. This behavior can also be seen in the real data.

Keywords: LASSO, LARS, multicolinearity

ABSTRAK

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

Oleh

PUTRI MAULIDINA FADILAH

LASSO merupakan metode yang dapat mengatasi masalah multikolinearitas dengan menyusutkan beberapa koefisien regresi mendekati nol dan bahkan menjadi tepat nol. Dengan algoritma LARS perhitungan LASSO menjadi lebih efisien dibandingkan pemrograman kuadratik. Tujuan dari penelitian ini adalah untuk mengetahui performa metode LASSO dalam mengatasi multikolinearitas dan membandingkan nilai dugaannya dengan Metode Kuadrat Terkecil. Hasil menunjukkan bahwa metode LASSO tidak dapat mengatasi multikolinearitas ketika seluruh variabel bebas berkorelasi tinggi. Namun, LASSO menangani multikolinearitas dengan baik ketika korelasi terjadi hanya pada sebagian variabel bebas. Pada jumlah data yang lebih besar, LASSO menghasilkan dugaan koefisien regresi yang lebih baik daripada MKT. Hal tersebut juga dapat terlihat pada data real.

Kata Kunci: LASSO, LARS, multikolinearitas

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

Oleh

PUTRI MAULIDINA FADILAH

Skripsi

Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA SAINS

pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2018

Judul Skripsi : **ANALISIS REGRESI DENGAN METODE *LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR* (LASSO) DALAM MENANGANI MULTIKOLINEARITAS**

Nama Mahasiswa : *Putri Maulidina Fadilah*

Nomor Pokok Mahasiswa : 1417031091

Program Studi : Matematika

Fakultas : Matematika dan Ilmu Pengetahuan Alam

MENYETUJUI

1. Komisi Pembimbing

Ir. Netti Herawati, M.Sc., Ph.D.
NIP. 19650125 199003 2 001

Drs. Tiryono Ruby, M.Sc., Ph.D.
NIP. 19620704 198803 1 002

2. Ketua Jurusan Matematika

Prof. Dra. Wamiliana, M.A., Ph.D.
NIP. 19631108 198902 2 001

MENGESAHKAN

1. Tim Penguji

Ketua : **Ir. Netti Herawati, M.Sc., Ph.D.** ....

Sekretaris : **Drs. Tiryono Ruby, M.Sc., Ph.D.** ....

Penguji
Bukan Pembimbing : **Drs. Eri Setiawan, M.Si.** ....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Prof. Warsito, S.Si., D.E.A., Ph.D.
NIP. 19710212 199512 1 001

Tanggal Lulus Ujian Skripsi : **14 Maret 2018**

PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan di bawah ini :

Nama	: **Putri Maulidina Fadilah**
Nomor Pokok Mahasiswa	: **1417031091**
Jurusan	: **Matematika**
Judul Skripsi	: **Analisis Regresi dengan Metode *Least Absolute Shrinkage and Selection Operator* (LASSO) dalam Menangani Multikolinearitas**

Dengan ini menyatakan bahwa penelitian ini adalah hasil pekerjaan saya sendiri dan apabila kemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, Maret 2018

Yang Menyatakan

METERAI TEMPEL

8A928ADC002842544

6000

ENAM RIBU RUPIAH

Putri Maulidina Fadilah

RIWAYAT HIDUP

Penulis bernama lengkap Putri Maulidina Fadilah, lahir di Bandar Lampung pada 5 Agustus 1996. Penulis merupakan anak kedua dari 3 bersaudara, pasangan bapak Hi. Syaifullah dan ibu Hj. Sukartini.

Penulis menempuh pendidikan dasar di SD Negeri 2 Rawalaut dari tahun 2002 – 2008. Kemudian melanjutkan pendidikan di SMP Negeri 2 Bandar Lampung dan lulus pada tahun 2011. Kemudian menempuh pendidikan di SMA Negeri 9 Bandar Lampung dan lulus pada tahun 2014.

Pada tahun 2014, penulis diterima sebagai mahasiswi di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung melalui jalur Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN). Pada tahun 2017 penulis melakukan Kerja Praktik di Kantor Otoritas Jasa Keuangan Lampung dan sebagai salah satu bentuk pengabdian kepada masyarakat penulis melaksanakan Kuliah Kerja Nyata di Desa Kedaloman, Kecamatan Gunung Alip, Kabupaten Tanggamus.

KATA INSPIRASI

"Boleh jadi, kamu membenci sesuatu, padahal ia amat baik bagimu, dan boleh jadi (pula) kamu menyukai sesuatu, padahal ia amat buruk bagimu. Allah yang paling mengetahui, sedangkan kamu tidak mengetahui."

(QS. Al-Baqarah : 216)

"*Hai orang-orang yang beriman, jadikanlah sabar dan shalat sebagai penolongmu, sesungguhnya Allah bersama orang-orang yang sabar*".

(QS. Al-Baqarah : 153)

"Barang siapa yang menempuh perjalanan untuk mencari ilmu, niscaya Allah subhanahu wata'ala menyediakan jalan untuknya menuju surga."

(H.R. Abu Daud, Tirmidzi, dan Ibnu Majjah)

PERSEMBAHAN

Karyaku yang sederhana ini kupersembahkan kepada:

Ayah dan Mamah

Terima kasih kepada Ayah dan Mamah yang selalu mendo'akan kesuksesanku, memberi semangat, nasihat, dukungan serta kasih sayang yang tiada henti.

Kakakku Opi dan Kakak Iparku Emma

Terima kasih kepada Aa' Opi dan Mba Emma yang selalu memberikan semangat, do'a, dan motivasi yang sangat berharga.

Adikku Nadia

Terima kasih kepada Adik yang selalu memberikan semangat dan keceriaan dalam hidupku.

Sahabat-sahabatku Tiara, Andan, Yutia, Intan, dan Susan

Terima kasih kepada para sahabatku yang selalu memberikan semangat, do'a, dan motivasi, serta kenangan indah selama ini.

Almamater dan Negeriku

SANWACANA

Puji syukur kehadirat Allah SWT yang telah memberikan rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi dengan judul “Analisis Regresi dengan Metode *Least Absolute Shrinkage and Selection Operator* (LASSO) dalam Menangani Multikolinearitas” dengan baik dan tepat pada waktunya.

Penulis menyadari bahwa skripsi ini dapat terselesaikan dengan baik karena dukungan, bimbingan, saran, serta do’a dari berbagai pihak. Oleh karena itu, dalam kesempatan ini penulis menyampaikan terima kasih kepada:

1. Ibu Ir. Netti Herawati, M.Sc., Ph.D., selaku dosen pembimbing satu yang telah memberikan bimbingan, pengarahan, dan saran kepada penulis dalam mengerjakan skripsi.
2. Bapak Drs. Tiryono Ruby, M.Sc., Ph.D., selaku pembimbing dua yang telah memberikan saran serta pembelajaran yang sangat bermanfaat dalam menyelesaikan skripsi.
3. Bapak Drs. Eri Setiawan, M.Si., selaku pembahas dan penguji skripsi yang telah memberikan evaluasi, arahan, dan saran demi perbaikan skripsi.
4. Ibu Widiarti, S.Si., M.Si., selaku dosen pembimbing akademik.

5. Ibu Prof. Dra. Wamiliana, M.A., Ph.D., selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
6. Seluruh dosen Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
7. Ayah dan Mamah tercinta yang selalu mendo'akan kesuksesan dunia dan akhirat.
8. Aa', mba Emma, dan Adik yang telah mendo'akan dan memberi saran serta keceriaan.
9. Sahabat-sahabat tersayang, Tiara, Andan, Yutia, Intan, Susan, Linda, Diba, Megita, Mega, Nia, Raafika yang telah mendo'akan, memberi dukungan dan kenangan indah kepada penulis.
10. Fietra, Mona, Ulfa, Jelli, Rium, dan teman-teman satu bimbingan lainnya, terima kasih atas semangat dan saran selama penyelesaian skripsi.
11. HIMATIKA yang telah memberikan pengalaman berharga.
12. Teman-teman mahasiswa Jurusan Matematika angkatan 2014.

Penulis menyadari bahwa dalam penulisan skripsi ini masih jauh dari sempurna, sehingga informasi tambahan, saran, dan kritik untuk pengembangan lebih lanjut sangat diharapkan. Semoga skripsi ini dapat bermanfaat bagi kita semua.

Bandar Lampung,   Maret 2018

Penulis

Putri Maulidina Fadilah

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Analisis regresi merupakan suatu analisis statistika yang paling banyak digunakan dalam aplikasinya. Tidak sedikit masalah-masalah spesifik muncul ketika dilakukan analisis. Salah satu masalah yang sering muncul yakni multikolinearitas.

Multikolinearitas merupakan masalah yang muncul ketika melakukan analisis regresi berganda dimana terdapat korelasi atau hubungan kuat antara dua atau lebih variabel bebas. Adanya multikolinearitas menyebabkan penduga kuadrat terkecil memiliki ragam yang lebih besar. Hal tersebut dapat berpengaruh pada tingkat akurasi prediksi model dan menyebabkan kesalahan dalam pengambilan keputusan. Menurut Hastie *et al*. (2015), akurasi prediksi terkadang dapat ditingkatkan dengan mengecilkan nilai koefisien regresi, atau mengatur beberapa koefisien menjadi nol.

Least Absolute Shrinkage and Selection Operator (LASSO) pertama kali dikenalkan oleh Tibshirani (1996) dan merupakan teknik regresi yang melakukan pendugaan dengan meminimumkan jumlah kuadrat galat dengan suatu kendala L_1, $\sum_{j=1}^{p}|\hat{\beta}_j| \leq t$ dimana *t* merupakan parameter tuning yang mengontrol besarnya penyusutan koefisien regresi dengan nilai $t \geq 0$. Karena kendala tersebut, LASSO dapat menyusutkan beberapa koefisien regresi bahkan menjadi nol sehingga LASSO juga dapat berfungsi sebagai seleksi variabel.

Berdasarkan hal-hal tersebut, performa metode LASSO akan diteliti menggunakan data simulasi kemudian diterapkan pada data riil yang telah terdeteksi adanya multikolinearitas.

1.2 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Mengetahui performa metode *Least Absolute Shrinkage and Selection Operator* (LASSO) dalam mengatasi multikolinearitas.
2. Membandingkan nilai dugaan menggunakan *Least Absolute Shrinkage and Selection Operator* (LASSO) dengan Metode Kuadrat Terkecil (MKT).

1.3 Manfaat Penelitian

Manfaat dari penelitian ini adalah menambah pengetahuan bagi penulis dan memberi masukkan kepada para peneliti dan pembaca tentang metode pendugaan LASSO untuk menganalisis data yang mengandung multikolinearitas.

II. TINJAUAN PUSTAKA

2.1 Analisis Regresi

Menurut Montgomery dan Peck (1992), analisis regresi merupakan suatu teknik untuk memeriksa dan memodelkan hubungan antar variabel. Dalam regresi sederhana dikaji dua variabel yaitu variabel terikat dan variabel bebas, sedangkan dalam regresi berganda melibatkan dua atau lebih variabel bebas. Dalam analisis regresi suatu persamaan regresi hendak ditentukan dan digunakan untuk menggambarkan pola atau bentuk fungsi hubungan yang terdapat antar variabel.

Persamaan regresi adalah persamaan matematik yang memungkinkan untuk meramalkan nilai-nilai suatu variabel terikat dari nilai-nilai satu atau lebih variabel bebas. Regresi diterapkan pada semua jenis peramalan, dan tidak harus berimplikasi suatu regresi mendekati nilai tengah populasi (Badri, 2012).

Model linier artinya linier dalam parameter (Draper & Smith 1992). Jika terdapat vektor input $\mathbf{x}^T = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p)$ dan digunakan untuk menduga luaran nilai Y yang berupa bilangan riil, maka model regresi linier memiliki bentuk sebagai berikut,

$$Y_i = \beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j + \varepsilon_i \tag{2.1}$$

dengan :

Y_i = Vektor variabel terikat berukuran nx1

β_0 = Intersep

x_{ij} = Matriks variabel bebas berukuran nx(p+1)

β_j = Slope atau kemiringan

ε_i = Vektor galat berukuran nx1

2.2 Regresi Linear Berganda

Analisis regresi linier berganda merupakan analisis hubungan secara linier antara dua atau lebih variabel bebas dengan satu variabel terikat. Analisis regresi linear berganda sebenarnya sama dengan analisis regresi linear sederhana, hanya variabel bebasnya lebih dari satu buah. Menurut Montgomery dan Runger (2011), misalkan n > k onservasi, dan x_{ij} dari pengamatan ke- i observasi. Model pengamatannya adalah

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \varepsilon_i \qquad (2.2)$$

dengan :

i = 1, 2, 3, ..., n

k = 1, 2, 3, ..., n

Y_i = Variabel terikat pengamatan ke-i

X_{ki} = Variabel bebas pengamatan ke-i

β_0 = Konstanta (parameter)

β_k = Koefisien regresi atau *slope* (parameter) ke-k

ε_i = Sisaan (galat) pengamatan ke-i

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

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \quad \text{dengan } \boldsymbol{\varepsilon} \sim N(0, \sigma^2 I) \tag{2.3}$$

dengan :

Y = Vektor n x 1 variabel terikat

X = Matriks n x k variabel bebas

β = Vektor k x 1 koefisien variabel bebas

£ = Vektor n x 1 variabel acak galat dengan E(ε) = 0 dan matriks ragam peragam $\sigma^2(\varepsilon) = \sigma^2 I$

2.3 Metode Kuadrat Terkecil (MKT)

Metode kuadrat terkecil merupakan metode yang digunakan untuk menduga koefisien regresi linear dengan cara meminimumkan jumlah kuadrat galat (Hastie, *et al*., 2008).

Menurut Sembiring (1995), untuk mencari nilai-nilai β yaitu dengan meminimumkan bentuk kuadrat

$$\begin{aligned} Q(\beta_j) &= (Y - X\beta)'(Y - X\beta) \\ &= Y'Y - 2Y'X\beta + \beta'X'X\beta \end{aligned} \tag{2.4}$$

kemudian dicari turunan dari $Q(\beta_j)$ secara parsial terhadap β_j; $j = 1,2,\ldots,k$ dan disamakan dengan nol.

$$\begin{bmatrix} n & \sum_{i=1}^{n} X_{i1} & \sum_{i=1}^{n} X_{i2} & \dots & \sum_{i=1}^{n} X_{ik} \\ \sum_{i=1}^{n} X_{i1} & \sum_{i=1}^{n} X_{i1}^2 & \sum_{i=1}^{n} X_{i2}X_{i1} & \dots & \sum_{i=1}^{n} X_{ik}X_{i1} \\ \sum_{i=1}^{n} X_{i2} & \sum_{i=1}^{n} X_{i1}X_{i2} & \sum_{i=1}^{n} X_{i2}^2 & \dots & \sum_{i=1}^{n} X_{ik}X_{i2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} X_{ik} & \sum_{i=1}^{n} X_{i1}X_{ik} & \sum_{i=1}^{n} X_{i2}X_{ik} & \dots & \sum_{i=1}^{n} X_{ik}^2 \end{bmatrix} \begin{bmatrix} b_o \\ b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_{11} & X_{21} & \dots & X_{n1} \\ X_{12} & X_{22} & \dots & X_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{ik} & X_{2k} & \dots & X_{nk} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix}$$

Atau secara lengkap jika ditulis kedalam bentuk matriks menjadi

$$X^T X\hat{\beta} = X^T Y \qquad (2.5)$$

Bila X’X tidak singular maka ada inversnya, sehingga diperoleh penduga untuk MKT

$$\hat{\beta} = (X^T X)^{-1} X^T Y \qquad (2.6)$$

Sifat-sifat penduga metode kuadrat terkecil adalah sebagai berikut

1. $\hat{\beta}$ linear

   $\hat{\beta}$ linear jika $\hat{\beta}$ merupakan fungsi linear dari β

$$\begin{aligned} \hat{\beta} &= (X^T X)^{-1} X^T Y \\ &= (X^T X)^{-1} X^T (X\beta + \varepsilon) \\ &= (X^T X)^{-1} X^T X\beta + (X^T X)^{-1} X^T \varepsilon \\ &= I\beta + (X^T X)^{-1} X^T \varepsilon \end{aligned}$$

2. $\hat{\beta}$ tak bias

$\hat{\beta}$ adalah penduga tak bias jika $E(\hat{\beta}) = \beta$

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

$$= E((X^TX)^{-1}X^TX\beta + \varepsilon)$$

$$= E((X^TX)^{-1}X^TX\beta + (X^TX)^{-1}X^T\varepsilon)$$

$$= (X^TX)^{-1}X^TX\beta + (X^TX)^{-1}X^TE(\varepsilon)$$

$$= (X^TX)^{-1}X^TX\beta$$

$$= \beta$$

Sehingga $\hat{\beta}$ merupakan penduga tak bias dari β

3. $\hat{\beta}$ memiliki variansi minimum

$$var(\hat{\beta}) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T]$$

$$= E[((X^TX)^{-1}X^TY - \beta)((X^TX)^{-1}X^TY - \beta)^T]$$

$$= E[\ ((X^TX)^{-1}X^T(X\beta + \varepsilon) - \beta)((X^TX)^{-1}X^T(X\beta + \varepsilon) - \beta)^T]$$

$$= E[((X^TX)^{-1}X^TX\beta + (X^TX)^{-1}X^T\varepsilon) - \beta)((X^TX)^{-1}X^TX\beta +$$
$$(X^TX)^{-1}X^T\varepsilon) - \beta)^T]$$

$$=E[(\beta + (X^TX)^{-1}X^T\varepsilon - \beta)(\beta + (X^TX)^{-1}X^T\varepsilon - \beta)^T]$$

$$=E[((X^TX)^{-1}X^T\varepsilon)((X^TX)^{-1}X^T\varepsilon)^T]$$

$$=E[((X^TX)^{-1}X^T\varepsilon)((X^TX)^{-1}X^T\varepsilon)^T]$$

$$=E[((X^TX)^{-1}X^T\varepsilon^T\varepsilon\ X\ (X^TX)^{-1}]$$

$$=E[((X^TX)^{-1}X^T\ E(\varepsilon\varepsilon^T)\ X\ (X^TX)^{-1}]$$

$$=(X^TX)^{-1}X^TX\ (X^TX)^{-1}\ E[\varepsilon\varepsilon^T]$$

$$=(X^TX)^{-1}\sigma^2$$

$Var(\hat{\beta}) = (X^T X)^{-1}\sigma^2$ merupakan varians terkecil dari semua penaksir linear tak bias.

Estimator kuadrat terkecil yang memenuhi sifat linear, tak bias, dan mempunyai variansi minimum ini bersifat *Best Linear Unbiased Estimator* (*BLUE)*

2.4 Multikolinearitas

Syarat multikolinearitas pertama kali dikemukakan 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 sebagai berikut.

$$VIF_{(j)} = \frac{1}{(1-R_j^2)} \quad j = 1, 2, \ldots, k \qquad (2.7)$$

R_j^2 merupakan koefisien determinasi yang didapat dari variabel prediktor X_j yang diregresikan dengan variabel prediktor lainnya. Jika nilai VIF(j) lebih besar dari 10 maka terjadi masalah multikolinearitas.

2.4.1 Dampak Multikolinearitas

Dampak dari multikolinearitas antara lain :

1. Koefisien Partial Regresi tidak terukur secara presisi. Oleh karena itu nilai standar errornya besar.
2. Perubahan kecil pada data dari sampel ke sampel akan menyebabkan perubahan drastis pada nilai koefisien regresi partial.
3. Perubahan pada satu variabel dapat menyebabkan perubahan besar pada nilai koefisien regresi parsial variabel lainnya.
4. Nilai selang kepercayaan sangat lebar, sehingga akan menjadi sangat sulit untuk menolak hipotesis nol pada sebuah penelitian jika dalam penelitian tersebut terdapat multikolinearitas

Menurut Jollife (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 dengan galat menjadi tepat pada nol atau mendekati nol (Tibshirani, 1996).

Persamaan secara umum LASSO dinyatakan sebagai berikut:

$$\mathbf{Y}^{**} = \mathbf{X}^{**}\boldsymbol{\beta} + \boldsymbol{\varepsilon}^{**} \qquad (2.8)$$

dengan :

Y** = Vektor variabel terikat berukuran (n x 1)

X** = Matriks variabel bebas berukuran (n x p)

β = Vektor dari koefisien LASSO berukuran (k+1) x 1

ε** = Vektor galat berukuran (n x 1)

Menurut Zhao dan Yu (2006), estimasi LASSO didefinisikan dengan :

$$\hat{\beta}^{LASSO} = \underset{\beta}{\operatorname{argmin}} \|Y - \boldsymbol{X}\beta\|_2^2 + t\,\|\beta\|_1 \tag{2.9}$$

Dengan syarat $\|\beta\|_1 \leq t$, dimana *t* merupakan parameter tuning yang mengontrol penyusutan koefisien LASSO dengan $t \geq 0$. Jika $t < t_0$ dengan $t_0 = \|\hat{\beta}_j\|_1$ maka akan menyebabkan koefisien menyusut mendekati 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 (Tibshirani, 1996).

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

Untuk mendapatkan solusi penduga LASSO tidak dapat diperoleh dalam bentuk tertutup, tetapi harus menggunakan pemrograman kuadratik (Tibshirani, 1996).

Setelah publikasi pertama tahun 1996, makalah LASSO tidak mendapatkan perhatian sampai tahun 2002 setelah berkembangkannya algoritma LAR (*Least Angle Regresion*) oleh Hastie. Modifikasi dari LAR untuk LASSO menghasilkan efisiensi algoritma dalam menduga solusi penduga koefisien LASSO dengan komputasi yang lebih cepat dibandingkan pemrograman kuadratik (Soleh dan Aunuddin, 2013). Menurut Hastie, *et al*. (2008) terdapat algoritma yang efisien untuk menghitung solusi model penuh dengan *t* yang beragam.

2.6 *Least Absolute Regression* (LAR)

Least Absolute Regression (LAR) merupakan algoritma yang lebih efisien digunakan karena LAR mempunyai modifikasi untuk mempermudah dalam komputasi LASSO dibandingkan pemrograman kuadratik.

Menurut Efron, *et al*. (2004), LAR melakukan estimasi $\hat{\boldsymbol{\mu}} = \boldsymbol{X}^{*}\hat{\boldsymbol{\beta}}$, dengan langkah-langkah yang berurutan dan di setiap langkah akan menambah satu kovarian ke dalam model. Nilai $\hat{\boldsymbol{\mu}}$ diperoleh dari teknik iterasi dengan nilai awal $\hat{\boldsymbol{\mu}}_0 = \mathbf{0}$.

Berikut adalah langkah-langkah estimasi koefisien LASSO dengan algoritman LAR.

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

$$\hat{\boldsymbol{c}} = \boldsymbol{X}'(\boldsymbol{y} - \hat{\boldsymbol{\mu}}) \qquad (2.11)$$

2. Menentukan korelasi saat mutlak terbesar

$$\hat{C} = \max\{|\hat{c}_j|\} \tag{2.12}$$

Maka diperoleh s_j = sign $\{\hat{c}_j\}$ untuk j є A

3. Menentukan X_A, dimana himpunan A merupakan himpunan indeks aktif dari variabel prediktor {1, 2, …, m}. Himpunan indeks aktif A ditentukan berdasarkan nilai korelasi mutlak terbesar, didefinisikan matriks berikut.

$$\mathbf{X_A} = (\ldots\ s_j\ \boldsymbol{X_j^*}\ \ldots)\ ,\ j \in A \tag{2.13}$$

dengan tanda s_j bernilai ±1, maka

$$\mathbf{G_A} = \boldsymbol{X_A'}\mathbf{X_A} \text{ dan } A_A = (\mathbf{1}_A'\boldsymbol{G_A^{-1}}\mathbf{1}_A)^{-\frac{1}{2}} \tag{2.14}$$

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°. Nilai vektor *equiangular* dicari menggunakan rumus berikut.

$$\mathbf{u_A} = \mathbf{X_A}\ \mathbf{w_A} \text{ dengan } \mathbf{w_A} = A_A\ \boldsymbol{G_A^{-1}}\mathbf{1}_A \tag{2.15}$$

5. Menghitung *inner product* $\mathbf{a} \equiv \boldsymbol{X_A'}\boldsymbol{u_A}$ (2.16)

6. Menghitung vektor prediksi $\boldsymbol{\hat{\mu}_{A+}} = \boldsymbol{\hat{\mu}_A} + \boldsymbol{\hat{\gamma}}\boldsymbol{u_A}$ (2.17)

$$\hat{\gamma} = \min\nolimits^{+}_{j\in A^c}\left\{\frac{\hat{C}-\hat{c}}{A_A-a_j}, \frac{\hat{C}+\hat{c}}{A_A+a_j}\right\} \tag{2.18}$$

$\min^{+}_{j\in A^c}$ 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 $\hat{\gamma} = \frac{\hat{c}_m}{A_m}$.

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

$$sign(\hat{\beta}_j) = sign(\hat{\boldsymbol{c}}_j) = s_j \tag{2.19}$$

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

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

$\beta_j(\gamma)$ 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, $c_j(\gamma)$ tidak berubah tanda dan proses LAR berhenti dan menghapus j dari perhitungan vektor *equiangular* selanjutnya, dan variabel j dimasukkan kembali pada tahap perhitungan LAR selanjutnya, maka

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

2.7 *Mean Square Error* (MSE)

Jika $\hat{\beta}$ adalah penduga yang tak bias dari β, maka E $((\hat{\beta}) - \beta)^2$ sama dengan ragam penduga $\hat{\beta}$. Tetapi jika suatu $\hat{\beta}$ adalah penduga yang bias dari β, maka E $((\hat{\beta})-\beta)^2$ disebut dengan MSE dari $\hat{\beta}$.

MSE $(\hat{\beta})$ = E $((\hat{\beta}) - \beta)^2$

Bukti:

$$\begin{aligned} \text{MSE } (\hat{\beta}) &= \text{E } (\hat{\beta}^2 - 2\hat{\beta}\,\beta + \beta^2) \\ &= \text{E } (\hat{\beta}^2) - 2\text{E}(\hat{\beta})\,\beta + \beta^2 \\ &= \{ \text{ E } (\hat{\beta}^2) - 2\text{E}(\hat{\beta})\,\beta + \beta^2\} + \{(\text{E } (\hat{\beta}))^2 - (\text{E } (\hat{\beta}))^2\} \end{aligned}$$

$= \{ E(\hat{\beta}^2) - (E(\hat{\beta}))^2 \} + \{(E(\hat{\beta}))^2 - 2E(\hat{\beta})\beta + \beta^2\}$

$= \{ E(\hat{\beta}^2) - (E(\hat{\beta}))^2 \} + \{(E(\hat{\beta}) - \beta)^2\}$

$= \text{var}(\hat{\beta}) + (\text{bias}(\hat{\beta}))^2$

2.8 Validasi Silang

Validasi silang merupakan metode yang dapat digunakan untuk mengestimasi galat prediksi dalam meningkatkan ketepatan dari pemilihan model (James, *et al*., 2013). Validasi silang membagi data menjadi dua bagian, yaitu data *training* dan data *test*. Data *training* digunakan untuk mengepas nilai $\hat{\beta}$, sedangkan data *test* digunakan untuk menguji kebaikan prediksi dari $X\hat{\beta}$. Nilai validasi silang yang diperoleh merupakan penduga bagi sisaan prediksi (Izenman, 2008). Menurut Efron dan Tibshirani (1993), salah satu metode tipe validasi silang adalah *k-fold*. Dalam validasi silang *k-fold*, semua observasi dipartisi secara acak ke dalam k sub-contoh. Setiap sub-contoh digunakan sebagai data *training*. Proses validasi silang diulang sampai k kali dan setiap satu sub-contoh digunakan hanya sekali dalam data *test* (Soleh dan Aunuddin, 2013). Nilai galat prediksi ($\widehat{PE}$) diduga oleh CV dengan menggunakan persamaan berikut.

$$\widehat{PE} = CV = \frac{1}{k}\sum_{i=1}^{k}(y_i - \hat{y}_{-i(k)})^2 \qquad (2.20)$$

dimana $\hat{y}_{-i(k)}$ adalah dugaan y pada saat *fold* ke-k tidak digunakan dalam menduga model. Keuntungan menggunakan 5-*folds* atau 10-*folds* validasi silang karena akan menghasilkan ragam rendah (James, *et al*., 2013).

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilakukan pada semester ganjil tahun ajaran 2017/2018. Bertempat di jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.

3.2 Data Penelitian

Data yang digunakan dalam penelitian ini adalah data simulasi dan data sekunder dari situs resmi Bank Indonesia. Untuk data simulasi yakni data yang dibangkitkan dengan variabel bebas sebanyak p = 6 dan n observasi yaitu n = 25, 50, dan 75 yang diulang sebanyak 100 kali dengan $\beta_0 = 0$ dan $\beta_1 = \beta_2 = \ldots = \beta_p = 1$.

Untuk mendapatkan data kolinearitas pada setiap himpunan data, X_p dibangkitkan menggunakan simulasi Monte Carlo dengan persamaan sebagai berikut :

$$X_{ij} = (1-\rho^2)^{1/2} z_{ij} + \rho z_{i(p+1)} \qquad ; i = 1, 2, \ldots, n \quad j = 1, 2, \ldots, p \tag{3.1}$$

dimana $z_{i1}, z_{i2}, \ldots, z_{i(p+1)}$ merupakan data yang dibangkitkan dalam bentuk normal standar atau berdistribusi normal N(0, 1) dan ρ ditentukan sehingga korelasi antarvariabel bebas diberikan oleh ρ^2. Dua himpunan dari variabel yang saling berkorelasi dalam penelitian ini dibuat berdasarkan nilai $\rho = 0.99$. Untuk lebih jelasnya dapat dilihat pada Tabel 3.1

Tabel 1. Simulasi Data X_{ip} Untuk 6 Variabel Bebas

Jumlah Variabel Bebas	Simulasi Data x_{ip}
6	$Z_{ij} - N(0, 1)$ *j* = 1, 2, …, 7 $X_{i1} = (1-\rho^2)^{1/2} z_{i1} + \rho z_{i7}$ $X_{i2} = (1-\rho^2)^{1/2} z_{i2} + \rho z_{i7}$ $X_{i3} = (1-\rho^2)^{1/2} z_{i3} + \rho z_{i7}$ $X_{i4} = (1-\rho^2)^{1/2} z_{i4} + \rho z_{i7}$ $X_{i5} = (1-\rho^2)^{1/2} z_{i5} + \rho z_{i7}$ $X_{i6} = (1-\rho^2)^{1/2} z_{i6} + \rho z_{i7}$

Variabel terikat (Y) untuk setiap p variabel bebas diperoleh berdasarkan model $Y = X\beta + \varepsilon$ dimana β adalah $\beta_{i,j} = 1$; untuk i = j , dan 0 selainnya. Dengan ε dibangkitkan berdasarkan distribusi normal N (0, 1) sehingga Y merupakan kombinasi linear dari *p* variabel bebas ditambah galat yang ditampilkan pada tabel berikut.

Tabel 2. Variabel Terikat (Y) Untuk 6 Variabel Bebas

p	Y
6	$Y = X_1 + X_2 + \cdots + X_p + N(0, 1) \qquad p = 1, 2, \ldots, 6$

Data sekunder diperoleh dari website resmi Bank Indonesia yaitu data Peredaran uang di Indonesia tahun 2014 - 2016 dengan variabel terikat (Y) yang digunakan adalah jumlah uang beredar dan variabel bebas (X) yang digunakan adalah aktifa dalam negeri bersih (X_1), tagihan bersih kepada pemerintah pusat (X_2), tagihan kepada perusahaan bukan keuangan BUMN (X_3), simpanan berjangka (X_4), jumlah tabungan (X_5), dan tabungan dalam rupiah (X_6).

3.3 Metode Penelitian

Penelitian ini dilakukan secara studi pustaka, yaitu dengan mempelajari buku-buku teks penunjang dan karya ilmiah yang disajikan dalam bentuk jurnal. Untuk mempermudah perhitungan dengan hasil yang akurat, penulis menggunakan software R versi 3.4.2.

Adapun langkah-langkah yang dilakukan adalah sebagai berikut:

1. Melakukan simulasi data.
2. Mengidentifikasi multikolinearitas dengan melihat nilai VIF (*Variance Inflation Factor).*

3. Melakukanan analisis regresi dengan metode LASSO menggunakan algoritma LARS dengan langkah – langkah sebagai berikut :
   - Melakukan *center and scale* data.
   - Mencari vektor yang sebanding dengan vektor korelasi antara variabel prediktor dan galat dari setiap variabel prediktor pada seleksi pertama berdasarkan persamaan (2.11).
   - Mencari nilai mutlak korelasi tertinggi berdasarkan persamaan (2.12) dan mendapatkan nilai X_A.
   - Mencari nilai vektor *equiangular* dengan mencari terlebih dahulu nilai bobot dari variabel yang telah terseleksi dengan rumus $\mathbf{w_A} = A_A\, \boldsymbol{G_A^{-1}} \boldsymbol{1_A}$. Ukuran dan nilai vektor $\mathbf{w_A}$ akan berubah seiring dengan bertambahnya variabel yang terseleksi, maka vektor *equiangular* ($\mathbf{u_A}$) akan diperoleh.
   - Mencari vektor *inner product* dari masing-masing variabel prediktor tanpa memperhatikan urutan indeks himpunan variabel aktif berdasarkan persamaan (2.16).
   - Mencari vektor prediksi $\hat{\boldsymbol{\mu}}_{A+}$ dengan mencari dahulu nilai gamma dari variabel yang sudah terpilih berdasarkan persamaan (2.18).
   - Mencari nilai β yang merupakan calon koefisien LASSO.
   - Melakukan pengecekan $sign(\hat{\beta}_j) = sign(\hat{\boldsymbol{c}}_{\boldsymbol{j}}) = s_j$.
   - Mengulang langkah yang sama untuk setiap seleksi variabelnya hingga semua variabel prediktor telah terseleksi.
4. Mencari model terbaik dengan menggunakan validasi silang.
5. Menghitung MSE *(Mean Square Error)* dan AMSE (*Average Mean Square Error*) dari penduga β pada metode LASSO dan MKT dengan rumus :

$$\mathrm{MSE}(\hat{\beta}) = \frac{1}{m}\sum_{j=1}^{m}\left(\hat{\beta}_j - \beta_i\right)^2 \tag{3.2}$$

$$\mathrm{AMSE}(\hat{\beta}) = \frac{1}{m}\sum_{j=1}^{m}\left\|\hat{\beta}_j - \beta_i\right\|^2 \tag{3.3}$$

dimana m adalah banyaknya pengulangan, j= 1, 2, …, m, dan i= 0, 1, …, p.

6. Membandingkan MSE *(Mean Square Error)* dan AMSE (*Average Mean Square Error*) dari penduga β metode LASSO dengan penduga β MKT.

V. KESIMPULAN

Dari hasil-hasil yang telah didapat, kesimpulan tentang analisis regresi dengan metode *Least Absolute Shrinkage and Selection Operator* (LASSO) dapat dikemukakan sebagai berikut.

1. Metode LASSO baik digunakan ketika tingkat multikolinearitas hanya terdapat pada sebagian variabel bebas. Bila multikolinearitas terjadi pada seluruh variabel bebas, maka LASSO tidak baik untuk digunakan.
2. Berdasarkan nilai AMSE, model regresi yang diperoleh dari metode LASSO lebih baik untuk menduga variabel terikat daripada model MKT pada data yang mengandung multikolinearitas.
3. Model regresi yang diperoleh dari metode LASSO akan semakin baik pada jumlah data yang semakin besar. Hal tersebut terlihat dari nilai AMSE yang semakin kecil ketika jumlah data semakin banyak.
4. Pada analisis menggunakan data peredaran uang di Indonesia tahun 2014-2016, didapatkan model regresi terbaik dengan metode LASSO menggunakan algoritma LARS yaitu $\hat{Y} = 0.5960287\ X_1 + 0.9231144\ X_3 + 0.7628798\ X_4 + 0.5098366\ X_5$. Meskipun masalah multikolinearitas tidak teratasi pada data jumlah uang beredar, tetapi model yang didapatkan masih lebih baik untuk menduga jumlah uang

beredar di Indonesia daripada model yang didapatkan dari MKT. Dimana berdasarkan model yang didapat, jumlah uang beredar dipengaruhi oleh aktifa dalam negeri bersih (X_1), tagihan kepada perusahaan bukan keuangan BUMN (X_3), simpanan berjangka (X_4) dan Jumlah Tabungan (X_5).

DAFTAR PUSTAKA

Andana, A. P. 2017. Model Regresi Menggunakan Least Absolute Shrinkage and Selection Operator (LASSO) pada Data Banyaknya Gizi Buruk Kabupaten/ Kota di Jawa Tengah. *Jurnal Gausssian*. **6**(1): 21-30.

Badri, S. 2012. *Metode Statistika Untuk Penelitian Kuantitatif*. Penerbit Ombak, Yogyakarta.

Bank Indonesia. Statistik Ekonomi dan Keuangan Indonesia (SEKI). Http://www.bi.go.id/seki/tabel/TABEL1_1.xls. Diakses pada 17 Januari 2018.

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

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

Efron, B., *et al.* 2004. Least Angle Regression. *The Annals of Statistics*. **32** (2): 407-499.

Efron, B. and Tibshirani,R.J. 1993. *An Introduction to the Bootstrap*. Chapman and Hall, London.

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

Hastie, T., Tibshirani, R., and Friedman, J. 2008. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Ed. ke-2. Springer, New York.

Izenman, A. J. 2008. *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*. Springer, New York.

James, G. *et al.* 2013. *An Introduction to Statistical Learning with Application in R*. Springer, New York.

Jolliffe, I. T. 2002. *Principal Component Analysis*. Ed. ke-2. Springer-Verlag, New York.

McDonald, G. C. and Galarneau, D. I. 1975. A Monte Carlo Evaluation of Some Ridge-Type Estimators. *Journal of the American Statistical Association*. **70** (350): 407-416.

Montgomery, D. C. and Runger, G. C. 2011. *Applied Statistics and Probability For Engineers*. Ed. ke-5. John Wiley &Sons, New York.

Montgomery, D. C. and Peck, A. E. 1992. *Introduction to Linear Regression Analysis*. A Wiley-Intersection Publication, New York.

Sembiring, R. K. 1995. *Analisis Regresi*. Penerbit ITB Bandung, Bandung.

Soleh, A. M. dan Aunuddin. 2013. LASSO: Solusi Alternatif Seleksi Peubah dan Penyusutan Koefisien Model Regresi. *Indonesian Jurnal of Statistics.* **18** (1): 21-27.

Tibshirani, R. 1996. Regression Shrinkage and Selection via LASSO. *Journal of the Royal Statistical Society*. Series B (Methodological). **58** (1): 267-288.

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