

# PENERAPAN ANALISIS LASSO DAN GROUP LASSO DALAM MENGIDENTIFIKASI FAKTOR-FAKTOR YANG BERHUBUNGAN DENGAN TUBERKULOSIS DI JAWA BARAT\*

Stephan Chen<sup>1‡</sup>, Khairil Anwar Notodiputro<sup>2</sup>, and Septian Rahardiantoro<sup>3</sup>

<sup>1</sup>Department of Statistics, IPB University, Indonesia, stephanchen97@gmail.com

<sup>2</sup>Department of Statistics, IPB University, Indonesia, khairil@apps.ipb.ac.id

<sup>3</sup>Department of Statistics, IPB University, Indonesia, rahardiantoro\_14@apps.ipb.ac.id

<sup>‡</sup>corresponding author

**Indonesian Journal of Statistics and Its Applications (eISSN:2599-0802)  
Vol 4 No 1 (2020), 39 - 54**

Copyright © 2020 Stephan Chen, Khairil Anwar Notodiputro, and Septian Rahardiantoro. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Abstract

Tuberculosis is the deadliest infectious disease in Indonesia, and West Java is a province with the largest number of tuberculosis cases in Indonesia. This research was conducted to identify variables and groups of variables that could explain the number of tuberculosis cases in West Java. The data used has many explanatory variables, and these variables form groups. LASSO and group LASSO analysis can be used for variables selection and handle data that has many explanatory variables, and group LASSO analysis can be used on data with grouped variables. The results of the LASSO analysis, variables that can explain the number of tuberculosis cases in West Java are the number of people with disabilities, the number of pharmacy staff, the number of malnourished people, the number of people working and the number of cities. According to the group LASSO analysis, the variables that can explain the number of tuberculosis cases in West Java are variables in the health and environmental groups. The government can focus on these factors if they want to reduce the number of tuberculosis cases in West Java.

**Keywords:** group LASSO, LASSO, tuberculosis, West Java.

---

\* Received Jul 2019; Accepted Nov 2019; Published online on Feb 2020

## 1. Pendahuluan

### 1.1 Latar Belakang

Tuberkulosis merupakan penyakit menular paling mematikan di Indonesia, dan ketiga dalam penyakit secara keseluruhan setelah penyakit jantung dan penyakit pernapasan akut (Andini, 2018). Tuberkulosis merupakan penyakit menular yang disebabkan oleh bakteri *Mycobacterium tuberculosis*. Menurut *World Health Organization* (WHO) pada tahun 2017, Indonesia termasuk dalam delapan negara dengan kasus penyakit tuberkulosis terbesar di dunia, bersama dengan India, China, Filipina, Pakistan, Nigeria, Bangladesh dan Afrika Selatan. Masih menurut WHO, penyakit tuberkulosis adalah salah satu dari 10 penyebab utama kematian di seluruh dunia. Pada 2017, 10 juta orang menderita penyakit tuberkulosis, dan 1.6 juta meninggal karena penyakit ini (termasuk 0.3 juta orang dengan HIV). Sedangkan provinsi Jawa Barat digunakan dalam penelitian ini karena menurut Badan Pusat Statistik (BPS), salah satu provinsi dengan jumlah kasus terbesar adalah Jawa Barat yang pada tahun 2016, terjadi sebanyak 57247 kasus.

Seluruh data yang digunakan dalam penelitian ini diambil dari publikasi BPS pada tahun 2017. Peubah yang diambil sebagai peubah penjelas adalah peubah yang berkaitan dengan kesehatan, ekonomi, sumber daya manusia (SDM) dan lingkungan. Faktor-faktor ini dipilih untuk diteliti karena menurut penelitian yang dilakukan oleh Sejati & Sofiana (2015), secara keseluruhan penelitian tentang tuberkulosis melibatkan faktor-faktor tersebut. Keseluruhan data yang terkumpul memiliki 27 amatan, yaitu kabupaten/kota di Jawa Barat, dengan 27 peubah penjelas yang terkait dengan kesehatan, ekonomi, SDM dan lingkungan. Peubah respon yang digunakan adalah jumlah kasus tuberkulosis di Jawa Barat tahun 2017 untuk masing-masing kabupaten/kota. Jika melihat perbandingan ukuran antara amatan dan jumlah peubah penjelas yang digunakan, maka dapat dikatakan bahwa data ini termasuk yang memiliki dimensi besar, karena jumlah peubah penjelasnya sudah menyamai jumlah amatannya.

Saat memodelkan hubungan antara satu kejadian dengan kejadian lainnya, teknik statistika yang lazim digunakan adalah analisis regresi. Analisis regresi akan menghadapi masalah jika jumlah peubah penjelasnya sangat banyak, sehingga diperlukan metode alternatif untuk menyelesaikan persoalan ini. Penelitian tentang pemodelan tuberkulosis pernah dilakukan oleh Pratama & Wulandari (2015) di Provinsi Jawa Barat menggunakan metode *Geographically Weighted Negative Binomial Regression*. Penelitian sebelumnya masih mengandung kelemahan antara lain, metode ini cukup rumit, dan sulit diterapkan untuk data yang memiliki banyak peubah prediktor, dan prediktor-prediktor tersebut membentuk kelompok-kelompok. Salah satu cara untuk menyelesaikan masalah ini di dalam literatur statistika dikenal metode regresi terpinalti, yaitu metode yang pendugaan parameternya didasarkan pada meminimuman jumlah kuadrat galat yang diberi pinalti (Hastie et al., 2015). Beberapa metode regresi terpinalti yang umumnya dikenal adalah regresi LASSO (*least absolute shrinkage and selection operator*), regresi *ridge* dan *elastic-net*. Perbedaan utama dari analisis LASSO dan *ridge* terletak pada pinalti yang diberikan, dimana pada LASSO, pinalti yang diberikan dikalikan dengan mutlak dari koefisien regresi, sedangkan pada *ridge* pinalti yang diberikan dikalikan dengan kuadrat koefisien regresi. Hal ini

mengakibatkan LASSO dapat menyusutkan nilai koefisien regresi menjadi tepat nol, sedangkan *ridge* hanya menyusutkan nilai koefisien regresi mendekati nol (Hastie *et al.*, 2015). Metode *elastic-net* merupakan metode yang menggabungkan pinalti dari LASSO dan *ridge*. Metode LASSO dipilih dalam penelitian ini karena sesuai dengan data yang berdimensi besar. Selain itu, LASSO juga dapat melakukan seleksi peubah dengan menyusutkan koefisien regresi menjadi nol. Metode lanjutan dari LASSO, yaitu *group* LASSO juga digunakan dalam penelitian ini. Hal ini karena peubah penjelas yang digunakan membentuk kelompok peubah, yaitu kesehatan, ekonomi, SDM, dan lingkungan.

LASSO merupakan teknik untuk menyeleksi peubah pada data yang berdimensi besar (Tibshirani, 1996). LASSO bekerja dengan menambahkan kendala pada metode kuadrat terkecil. Penelitian sebelumnya yang dilakukan oleh Prabowo *et al.* (2015) menerapkan metode LASSO pada data yang memiliki banyak peubah penjelas. Hasilnya, koefisien regresi yang dihasilkan dari metode LASSO lebih dapat melakukan seleksi peubah penjelas dibandingkan metode regresi biasa. Beberapa macam analisis LASSO antara lain adalah LASSO, *group* LASSO dan *fused* LASSO. Data yang digunakan adalah data kasus tuberkulosis di Jawa Barat dan memiliki kriteria dengan peubah penjelas yang banyak dan membentuk kelompok. Hal ini menyebabkan analisis yang digunakan adalah LASSO dan *group* LASSO. Sedangkan analisis *fused* LASSO cocok untuk digunakan pada data yang strukturnya memiliki urutan seperti pada data-data genetika pada bidang biologi (Hastie *et al.*, 2015). *Group* LASSO merupakan pengembangan dari LASSO, dan metode *group* LASSO dikembangkan dengan menambahkan kendala kelompok ke dalam metode LASSO. Penelitian yang sebelumnya pernah dilakukan oleh Yuan & Lin (2006), menunjukkan hasil bahwa *group* LASSO yang termasuk dalam keluarga LASSO, cukup baik untuk data berdimensi tinggi dan peubah penjelasnya membentuk kelompok.

Berdasarkan analisis korelasi, terjadi korelasi antar beberapa peubah penjelas, atau biasa disebut multikolinearitas. Hal ini tidak menjadi masalah pada analisis ini, karena metode LASSO juga dapat menangani kasus multikolinearitas (Dewi, 2010; Soleh & Aunuddin, 2013; Soleh *et al.*, 2015). Terdapat data hilang pada penelitian ini, yaitu data tuberkulosis pada Kabupaten Bandung tahun 2017. Data hilang ini akan ditangani dengan metode Yates, dengan bantuan data dari tahun 2015 dan 2016. Penelitian yang sebelumnya mengimplementasikan metode ini, adalah penelitian oleh Kinansi (2017). Hasilnya adalah metode Yates ini memberikan pendugaan data hilang yang baik.

## 1.2 Tujuan Penelitian

Penelitian ini bertujuan untuk mengidentifikasi faktor-faktor yang berhubungan dengan jumlah kasus tuberkulosis di Provinsi Jawa Barat, menggunakan analisis LASSO dan *group* LASSO. Data yang digunakan berasal dari publikasi Badan Pusat Statistik (BPS). Data yang digunakan meliputi data terkait kesehatan, ekonomi, SDM dan lingkungan pada tahun 2017. Selanjutnya, hasil penelitian ini diharapkan berguna untuk membuat solusi dalam mengurangi kasus tuberkulosis di Provinsi Jawa Barat.

## 2. Metodologi

### 2.1 Data

Data yang digunakan merupakan data sekunder yang berasal dari Badan Pusat Statistik (BPS). Data merupakan data publikasi pada tahun 2018 yang berisi statistik Provinsi Jawa Barat pada tahun 2017. Struktur data yang digunakan mencakup data jumlah kasus tuberkulosis, data terkait kesehatan, data terkait ekonomi, data terkait SDM (sumber daya manusia) dan data terkait lingkungan. Setiap baris amatan, merupakan kabupaten/kota di Provinsi Jawa Barat. Jumlah amatan kabupaten/kota yang ada di provinsi Jawa Barat ada sebanyak 27 amatan. Data yang digunakan pada penelitian ini memuat data hilang. Data hilang ini akan diduga dengan metode Yates. Secara lengkap, peubah-peubah yang digunakan tersaji dalam Tabel 1. Seluruh peubah merupakan peubah numerik. Sebelum di analisis, data yang akan digunakan, terlebih dahulu distandardisasi minimum-maksimum, sehingga tidak terjadi ketidakseimbangan saat dilakukan analisis.

### 2.2 Model

#### a. Analisis LASSO

Analisis LASSO didasarkan pada model:

$${}^{(st)}Y_{LASSO} = \beta_0 + \beta_{1,1}{}^{(st)}X_{1,1} + \beta_{1,2}{}^{(st)}X_{1,2} + \dots + \beta_{4,27}{}^{(st)}X_{4,27} + \varepsilon \quad (1)$$

dengan  ${}^{(st)}Y_{LASSO}$  merupakan jumlah kasus tuberkulosis yang telah distandardisasi berdasarkan analisis LASSO,  $\beta_0, \beta_{1,1}, \beta_{1,2}, \dots, \beta_{4,27}$  merupakan koefisien-koefisien yang saling berkaitan dengan masing-masing peubah penjelas,  ${}^{(st)}X_{1,1}, {}^{(st)}X_{1,2}, \dots, {}^{(st)}X_{4,27}$  merupakan peubah penjelas yang telah distandardisasi dan kelompoknya seperti tercantum pada Tabel 1, dan  $\varepsilon$  merupakan sisaan.

#### b. Analisis Group LASSO

Analisis *group* LASSO didasarkan pada model:

$${}^{(st)}Y_{GroupLASSO} = \beta_0 + \beta_{1,1}{}^{(st)}X_{1,1} + \beta_{1,2}{}^{(st)}X_{1,2} + \dots + \beta_{4,27}{}^{(st)}X_{4,27} + \varepsilon \quad (2)$$

dengan  ${}^{(st)}Y_{GroupLASSO}$  merupakan jumlah kasus tuberkulosis yang telah distandardisasi menurut analisis *group* LASSO,  $\beta_0, \beta_{1,1}, \beta_{1,2}, \dots, \beta_{4,27}$  merupakan koefisien-koefisien yang saling berkaitan dengan masing-masing peubah penjelas,  ${}^{(st)}X_{1,1}, {}^{(st)}X_{1,2}, \dots, {}^{(st)}X_{4,27}$  merupakan peubah penjelas yang telah distandardisasi dan kelompoknya seperti tercantum pada Tabel 1, dan  $\varepsilon$  merupakan sisaan. Peubah-peubah ini dikelompokkan menurut kategori kesehatan, ekonomi, SDM dan lingkungan seperti pada Tabel 1.

Tabel 1: Daftar peubah, kelompok peubah dan penjelasannya yang digunakan dalam penelitian.

Peubah	Kelompok Peubah	Nama Peubah	Satuan
Y	-	Jumlah kasus tuberkulosis	Kasus
X <sub>1</sub>	Kesehatan	Jumlah orang disabel	Orang
X <sub>2</sub>	Kesehatan	Jumlah rumah sakit	Unit
X <sub>3</sub>	Kesehatan	Jumlah puskesmas	Unit
X <sub>4</sub>	Kesehatan	Jumlah tenaga medis	Orang
X <sub>5</sub>	Kesehatan	Jumlah tenaga kefarmasian	Orang
X <sub>6</sub>	Kesehatan	Jumlah penderita gizi buruk	Orang
X <sub>7</sub>	Kesehatan	Jumlah kasus HIV/AIDS	Kasus
X <sub>8</sub>	Ekonomi	Jumlah gelandangan dan pengemis	Orang
X <sub>9</sub>	Ekonomi	PDRB	Miliar rupiah
X <sub>10</sub>	Ekonomi	Laju pertumbuhan PDRB	Indeks
X <sub>11</sub>	Ekonomi	Jumlah keluarga miskin	Keluarga
X <sub>12</sub>	SDM	Tuna susila	Orang
X <sub>13</sub>	SDM	Harapan lama sekolah (HLS)	Tahun
X <sub>14</sub>	SDM	Jumlah orang bekerja	Orang
X <sub>15</sub>	SDM	Jumlah orang pengangguran terbuka	Orang
X <sub>16</sub>	SDM	Jumlah angkatan kerja	Orang
X <sub>17</sub>	SDM	IPM	Indeks
X <sub>18</sub>	SDM	Jumlah anak terlantar	Orang
X <sub>19</sub>	SDM	Jumlah lansia	Orang
X <sub>20</sub>	SDM	Jumlah korban narkoba	Orang
X <sub>21</sub>	Lingkungan	Jumlah kecamatan	Unit
X <sub>22</sub>	Lingkungan	Jumlah pedesaan	Unit
X <sub>23</sub>	Lingkungan	Jumlah perkotaan	Unit
X <sub>24</sub>	Lingkungan	Timbunan sampah kota	M <sup>3</sup> /hari
X <sub>25</sub>	Lingkungan	Sampah terangkut ke TPA	M <sup>3</sup> /hari
X <sub>26</sub>	Lingkungan	Proyeksi jumlah penduduk	Orang
X <sub>27</sub>	Lingkungan	Luas area	Km <sup>2</sup>

### c. Ukuran Kebaikan Model

Setelah seluruh tahapan analisis LASSO dan *group* LASSO dilakukan, dan didapatkan model serta hasilnya, ukuran kebaikan kedua model ini juga diidentifikasi. Ukuran kebaikan yang digunakan adalah MSE (*Mean Squared Error*),  $R^2$  (*R-squared*) dan  $R^2_{adj}$  (*adjusted R-squared*). MSE yang merupakan nilai rata-rata kuadrat *error* atau sisaan, digunakan untuk membandingkan kemampuan antara analisis LASSO dan *group* LASSO untuk data pada kasus ini, yang merupakan data dengan banyak peubah penjelas dan peubah penjelas tersebut membentuk kelompok. Melalui ukuran ini, akan dapat disimpulkan model terbaik antara analisis LASSO dan *group* LASSO,

yaitu dengan nilai MSE yang lebih kecil.  $R^2$  merupakan persentase keragaman yang mengukur kemampuan peubah penjelas pada model dalam menjelaskan keragaman peubah respon. Semakin nilai  $R^2$  ini mendekati 1 atau 100%, maka model tersebut semakin baik.  $R^2$  ini digunakan untuk melihat seberapa jauh kedua model ini dapat menjelaskan peubah respon. Sedangkan  $R_{adj}^2$  merupakan nilai  $R^2$  dengan mempertimbangkan faktor bias yang dihasilkan dari banyaknya peubah penjelas dalam model. Semakin banyak peubah penjelas yang masuk ke dalam model, maka nilai  $R_{adj}^2$  akan semakin mengecil karena semakin besar bias yang dihasilkan. Sama seperti  $R^2$ , semakin  $R_{adj}^2$  mendekati 1 atau 100%, maka model tersebut semakin baik.

## 2.3 Prosedur Analisis Data

### a. Melakukan Pendugaan Data Hilang

Data hilang adalah data jumlah kasus tuberkulosis di Kabupaten Bandung, akan diduga menggunakan metode Yates. Pendugaan ini dilakukan dengan:

$$\hat{y}_{ij} = \frac{n \times y_{i.} + t \times y_{.j} - y_{..}}{(n - 1)(t - 1)} \quad (3)$$

dengan :

$\hat{y}_{ij}$  : prediksi data hilang

$y_{i.}$  : total amatan ke- $i$  ( $i=1,2,\dots,n$ ) yang memuat data yang hilang

$y_{.j}$  : total waktu ke- $j$  ( $j=1,2,\dots,t$ ) yang memuat data yang hilang

$y_{..}$  : total pengamatan tidak termasuk data hilang

$n$  : banyaknya amatan

$t$  : banyaknya waktu

### b. Melakukan Eksplorasi Data

Melakukan eksplorasi data untuk melihat fakta-fakta menarik, seperti sebaran wilayah kasus tuberkulosis dengan melakukan pemetaan dan dengan grafik-grafik. Eksplorasi ini dilakukan dengan teknik statistika deskriptif.

### c. Melakukan Standardisasi Data

Sebelum dilakukan analisis LASSO dan *group* LASSO, data terlebih dahulu distandardisasi menggunakan standardisasi minimum-maksimum. Hasil dari metode ini akan membuat nilai seluruh data memiliki nilai minimum 0 dan maksimum 100.

### d. Melakukan Analisis LASSO

Data dianalisis menggunakan analisis LASSO dengan menentukan koefisien dari model 1. Analisis dilakukan dengan menggunakan *software* R 3.4.1 dengan paket *glmnet*. Langkah-langkah yang dilakukan adalah sebagai berikut:

- Menentukan lambda terbaik dengan validasi silang 10-*fold*
- Melakukan analisis untuk melihat koefisien masing-masing kelompok peubah penjelas
- Melakukan interpretasi

### e. Melakukan Analisis Group LASSO

Data kemudian dianalisis menggunakan analisis *group* LASSO dengan menentukan koefisien dari model 2. Analisis dilakukan dengan menggunakan *software* R 3.4.1 dengan paket `gglasso`. Langkah-langkah yang dilakukan adalah sebagai berikut:

- Menentukan lambda terbaik dengan validasi silang *10-fold*
- Melakukan analisis untuk melihat koefisien masing-masing kelompok peubah penjelas
- Melakukan interpretasi per peubah penjelas
- Melakukan analisis untuk melihat koefisien masing-masing peubah penjelas
- Melakukan interpretasi per kelompok peubah

### f. Mengidentifikasi Ukuran Keباikan Analisis LASSO dan Group LASSO

Hasil analisis yang telah didapat dari metode LASSO dan *group* LASSO kemudian dilihat nilai MSE, *R-squared* dan *adjusted R-squared* nya, dan dilihat seberapa jauh kebaikan dari kedua model ini.

## 3. Hasil dan Pembahasan

### 3.1 Penanganan Data Hilang

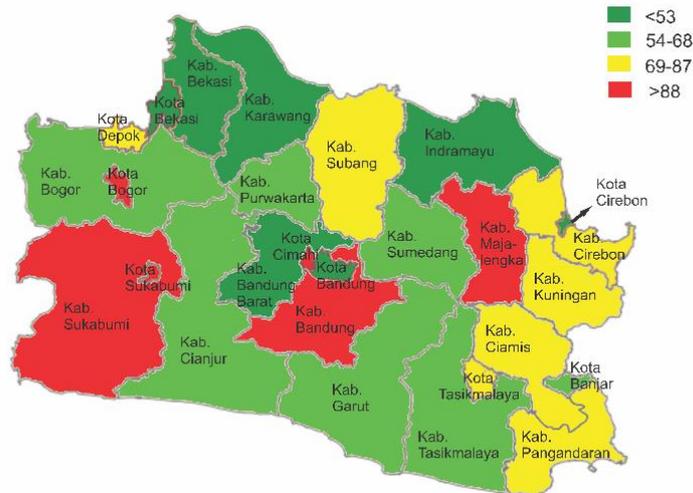
Data hilang ditangani dengan menggunakan metode Yates menggunakan persamaan (3). Setelah dilakukan perhitungan:

$$Y'_{2017KabBandung} = \frac{27 \times 10216 + 3 \times 28129 - 144822}{(27 - 1) \times (3 - 1)} = 4142.25$$

Didapat prediksi data hilang untuk jumlah kasus tuberkulosis di Kabupaten Bandung tahun 2017 menurut metode Yates adalah 4142.25.

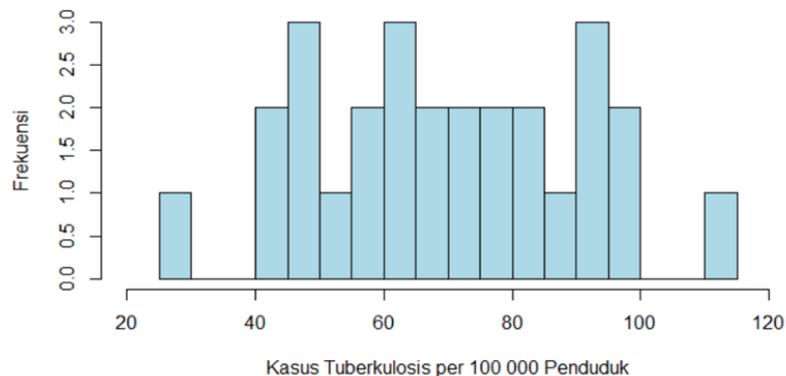
### 3.2 Eksplorasi Data

Sebaran kasus tuberkulosis per 100 000 penduduk di Jawa Barat pada tahun 2017 disajikan pada Gambar 1. Batasan warna dalam gambar, didapat dengan menggunakan nilai kuartil 1, median dan kuartil 3 sebagai pembatas. Kasus tuberkulosis dengan nilai antara median dan kuartil 3 (warna kuning) mayoritas terdapat pada kabupaten/kota disisi timur Jawa Barat, namun ada juga yang berada di sisi lainnya. Kasus tuberkulosis dengan nilai di bawah kuartil 1 (warna hijau tua), mayoritas berada pada sisi utara Jawa Barat. Kasus tuberkulosis dengan nilai antara kuartil 1 dan median (warna hijau muda) cukup menyebar merata dari selatan menuju tengah Jawa Barat. Sedangkan kasus tuberkulosis dengan nilai lebih dari kuartil 3 (warna merah), cukup acak tersebar di Jawa Barat.



Gambar 1: Jumlah kasus tuberkulosis per 100 000 penduduk di Jawa Barat tahun 2017.

Berdasarkan histogram pada Gambar 2, jumlah kasus tuberkulosis di Jawa Barat per 100 000 penduduk mayoritas berada pada kelas dengan titik tengah 59.14, 67.46 dan 92.44 dengan 4 kabupaten/kota. Data memiliki bentuk cenderung normal berdasarkan bentuk dari histogram.



Gambar 2: Histogram jumlah kasus tuberkulosis per 100 000 penduduk di Jawa Barat tahun 2017.

### 3.3 Analisis LASSO

Setelah dilakukan analisis LASSO, didapatkan hasil beberapa peubah yang mempengaruhi jumlah kasus tuberkulosis di Provinsi Jawa Barat. Hal ini terlihat dari koefisien peubahnya. Peubah yang memiliki koefisien tidak nol adalah peubah yang berpengaruh terhadap jumlah kasus tuberkulosis di Provinsi Jawa Barat. Berdasarkan hasil validasi silang 10-fold, didapatkan lambda ( $\lambda$ ) sebesar 3.42. Model yang didapat dengan analisis LASSO ini memiliki nilai MSE sebesar 116.34, *R-squared* sebesar

81.79% dan *adjusted R-squared* sebesar 77.45 %. Nilai koefisien dari masing-masing peubah penjelas dalam analisis LASSO ini dapat dilihat pada Tabel 2.

Tabel 2: Nilai koefisien peubah penjelas yang masuk ke dalam model LASSO.

Peubah	Koefisien
(Intersep)	0.83
$^{(st)}X_1$	0.08
$^{(st)}X_5$	0.04
$^{(st)}X_6$	0.03
$^{(st)}X_{14}$	0.57
$^{(st)}X_{23}$	0.15

Keterangan: Peubah yang tidak masuk dalam tabel memiliki koefisien sama dengan nol.

Dari model LASSO yang telah didapatkan, diketahui bahwa faktor-faktor yang dapat menjelaskan jumlah kasus tuberkulosis di Provinsi Jawa Barat menurut analisis LASSO adalah  $X_1$ ,  $X_5$ ,  $X_6$ ,  $X_{14}$  dan  $X_{23}$ . Faktor-faktor tersebut adalah jumlah orang difabel, jumlah tenaga kefarmasian, jumlah penderita gizi buruk, jumlah orang bekerja dan jumlah perkotaan. Seluruh peubah yang dapat menjelaskan jumlah kasus tuberkulosis di Jawa Barat memiliki koefisien positif dengan jumlah kasus tuberkulosis. Hal ini menunjukkan jika semakin tinggi nilai peubah penjelas tersebut, akan meningkatkan jumlah kasus tuberkulosis di Jawa Barat. Berdasarkan hasil ini, berikut adalah beberapa solusi untuk mengurangi kasus tuberkulosis. Pemerintah harus memperhatikan penderita disabilitas dan gizi buruk, karena memiliki nilai koefisien yang positif. Pemerintah dapat membuat program untuk mengurangi penderita gizi buruk seperti program pemberian makanan bergizi untuk anak kecil, pemberian vitamin rutin, dan program-program lainnya yang berkaitan dengan gizi. Penderita disabilitas juga harus diperhatikan kesehatannya, dengan membuat program-program kesehatan seperti pengecekan rutin, untuk para penderita disabilitas.

Jumlah orang bekerja juga memiliki koefisien yang positif dengan kasus tuberkulosis. Oleh karena itu, kemungkinan yang harus dilakukan adalah memperbaiki lingkungan tempat bekerja, agar para pekerja tidak mudah terkena kasus tuberkulosis, dan karena penyakit tuberkulosis ini menular, perusahaan dapat melakukan pengecekan rutin terhadap karyawannya. Kawasan perkotaan yang memiliki koefisien positif, juga menjadi perhatian. Pemerintah dapat memperbaiki kualitas lingkungan kawasan perkotaan, karena semakin banyak kota akan meningkatkan kasus tuberkulosis, yang menunjukkan bahwa kawasan perkotaan memiliki lingkungan yang kurang sehat. Kondisi polusi kemungkinan menjadi penyebabnya. Polusi tentu dapat mengurangi kekebalan imun manusia, sehingga akan lebih mudah terserang penyakit, terutama penyakit pernafasan seperti tuberkulosis ini. Meningkatkan kualitas wilayah perkotaan dapat dengan meningkatkan ruang terbuka hijau agar dapat mengurangi polusi udara. Hal yang tidak sesuai dengan substansi terjadi pada kasus tenaga kefarmasian. Seharusnya tenaga kefarmasian ini memiliki koefisien yang negatif terhadap jumlah kasus tuberkulosis. Hal ini juga menunjukkan bahwa tenaga

kefarmasian belum mampu mengurangi kasus tuberkulosis di Jawa Barat, mungkin saja tenaga kefarmasian perlu ditinjau kembali kualitasnya.

### 3.4 Analisis Group LASSO

#### a. Analisis per peubah

Seperti analisis LASSO, pada analisis *group* LASSO, peubah penjelas yang memiliki koefisien tidak nol, adalah peubah yang dapat menjelaskan jumlah kasus tuberkulosis di Provinsi Jawa Barat. Berdasarkan hasil validasi silang 10-fold, didapatkan lambda ( $\lambda$ ) sebesar 32.34. Model yang didapat dengan analisis *group* LASSO ini memiliki nilai MSE sebesar 88.17, *R-squared* sebesar 85.27% dan *adjusted R-squared* sebesar 68.09 %. Nilai koefisien dari masing-masing peubah penjelas, dalam analisis *group* LASSO dapat dilihat pada Tabel 3.

Hasil tersebut menunjukkan bahwa faktor-faktor yang dapat menjelaskan jumlah kasus tuberkulosis di Provinsi Jawa Barat menurut analisis *group* LASSO adalah seluruh peubah penjelas dalam kelompok kesehatan dan lingkungan. Peubah itu adalah jumlah orang difabel, jumlah rumah sakit, jumlah puskesmas, jumlah tenaga medis, jumlah tenaga kefarmasian, jumlah penderita gizi buruk, jumlah penderita HIV/AIDS, jumlah kecamatan, jumlah perdesaan, jumlah perkotaan, timbunan sampah kota, sampah terangkut ke TPA, proyeksi banyak penduduk dan luas area. Peubah jumlah rumah sakit, jumlah puskesmas, jumlah kasus HIV/AIDS dan jumlah pedesaan memiliki koefisien negatif terhadap jumlah kasus tuberkulosis. Artinya semakin tinggi nilai peubah penjelas tersebut, akan menurunkan jumlah kasus tuberkulosis di Jawa Barat. Sedangkan sisanya memiliki koefisien positif terhadap jumlah kasus tuberkulosis, yang artinya semakin tinggi nilai peubah penjelas tersebut, akan meningkatkan jumlah kasus tuberkulosis di Jawa Barat.

Tabel 3: Nilai koefisien peubah penjelas yang masuk ke dalam model *group* LASSO.

Peubah	Koefisien
(Intersep)	-1.78
$(st)X_1$	0.09
$(st)X_2$	-0.08
$(st)X_3$	-0.01
$(st)X_4$	0.06
$(st)X_5$	0.08
$(st)X_6$	0.07
$(st)X_7$	-0.02
$(st)X_{21}$	0.05
$(st)X_{22}$	-0.08
$(st)X_{23}$	0.24
$(st)X_{24}$	0.14
$(st)X_{25}$	0.02
$(st)X_{26}$	0.28
$(st)X_{27}$	0.15

Keterangan: Peubah yang tidak masuk dalam tabel memiliki koefisien sama dengan nol.

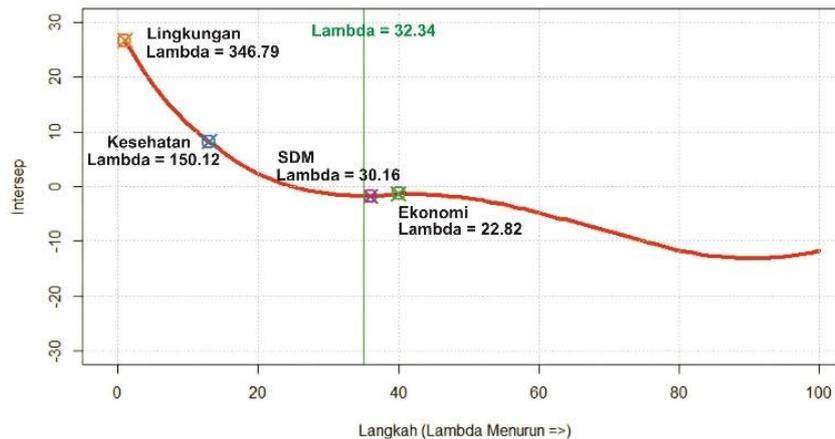
Berikut adalah beberapa solusi untuk mengurangi jumlah kasus tuberkulosis berdasarkan analisis *group* LASSO ini. Sama seperti sebelumnya, pemerintah harus memperhatikan orang difabel dan penderita gizi buruk karena memiliki koefisien positif. Pemerintah dapat membuat program untuk menekan angka gizi buruk. Pemerintah juga dapat menambah jumlah rumah sakit dan puskesmas yang memiliki koefisien negatif, untuk mengurangi jumlah kasus tuberkulosis. Rumah sakit dan puskesmas ini dapat meningkatkan pelayanan kesehatan bagi masyarakat. Jumlah kecamatan, proyeksi banyak penduduk dan luas area memiliki koefisien positif, sehingga semakin banyak kecamatan, berarti pelayanan kesehatan juga harus ditingkatkan sehingga lebih merata. Proyeksi banyak penduduk yang semakin besar juga berarti memiliki kemungkinan kasus tuberkulosis yang besar pula. Pada kabupaten/kota yang memiliki luas area yang besar, berarti harus diperhatikan pula pelayanan kesehatannya karena semakin sulit untuk pemeratakan pelayanan karena wilayah yang luas.

Jumlah kawasan perkotaan yang memiliki koefisien positif terhadap jumlah kasus tuberkulosis juga berarti wilayah perkotaan butuh perhatian ekstra pemerintah. Dapat dengan memperbanyak ruang terbuka hijau untuk mengurangi polusi udara. Sebaliknya, jumlah kawasan perdesaan yang memiliki hubungan negatif terhadap jumlah kasus tuberkulosis memiliki arti bahwa wilayah perdesaan lebih sehat dibandingkan wilayah perkotaan. Dimungkinkan karena wilayah perdesaan lebih kecil polusi lingkungannya dibandingkan wilayah perkotaan, sehingga wilayah perkotaan harus diatur agar polusinya paling tidak mirip dengan di desa. Timbunan sampah kota yang semakin besar, juga berpotensi meningkatkan jumlah kasus tuberkulosis, sehingga harus ada penanganan yang tepat, ini berkaitan dengan polusi yang dihasilkan oleh timbunan sampah ini.

Hal yang tidak sesuai dengan substansi terjadi pada peubah jumlah tenaga medis, jumlah tenaga kefarmasian dan sampah terangkut ke TPA yang memiliki koefisien positif dengan jumlah kasus tuberkulosis. Kemungkinan tenaga medis dan tenaga kefarmasian perlu ditinjau kembali kualitasnya, karena belum mampu mengurangi jumlah kasus tuberkulosis di Jawa Barat. Pada peubah jumlah penderita HIV/AIDS juga terjadi hal tidak sesuai dengan substansi karena memiliki koefisien negatif dengan jumlah kasus tuberkulosis.

#### **b. Analisis per kelompok peubah**

Analisis *group* LASSO juga dapat melihat besarnya pengaruh tiap kelompok peubah terhadap peubah respon. Hal tersebut dapat dilihat pada grafik yang disajikan dalam Gambar 3.



Gambar 3: Seluruh kemungkinan lambda ( $\lambda$ ) pada analisis *group* LASSO.

Analisis *group* LASSO juga dapat melihat besarnya pengaruh tiap kelompok peubah terhadap peubah respon yaitu jumlah kasus tuberkulosis di Jawa Barat. Hal tersebut dapat dilihat pada grafik yang disajikan dalam Gambar 3. Garis merah pada grafik di atas, menunjukkan seluruh kemungkinan lambda. Sumbu (x) merupakan langkah, yaitu indeks untuk masing-masing kemungkinan lambda, ditampilkan dari 0 sampai 100, dimana seiring bertambahnya indeks langkah ini, nilai lambda akan semakin menurun. Sumbu (y) merupakan nilai intersep pada model, saat langkah ke- $x$ . Semakin besar nilai lambda, maka akan semakin banyak peubah penjelas yang tidak masuk dalam model.

Berdasarkan Gambar 3, jika nilai lambda  $150.12 < \lambda \leq 346.79$ , maka hanya kelompok peubah lingkungan yang masuk ke dalam model. Jika nilai lambda  $30.16 < \lambda \leq 150.12$ , maka kelompok peubah lingkungan dan kesehatan akan masuk ke dalam model. Jika nilai lambda  $22.82 < \lambda \leq 30.16$ , maka kelompok peubah lingkungan, kesehatan dan SDM akan masuk ke dalam model. Terakhir, jika nilai lambda  $\lambda \leq 22.82$ , maka seluruh kelompok peubah akan masuk ke dalam model.

Berdasarkan hasil validasi silang *10-fold*, nilai lambda terbaik yang meminimumkan nilai rata-rata MSE, dan digunakan adalah 32.34. Hal ini mengakibatkan kelompok peubah SDM dan ekonomi tidak masuk dalam model *group* LASSO. Hasil ini juga menunjukkan bahwa urutan kelompok peubah yang paling penting sampai yang tidak penting, terhadap jumlah kasus tuberkulosis di Jawa Barat adalah lingkungan, kesehatan, SDM dan ekonomi.

Jika pemerintah ingin mengurangi jumlah kasus tuberkulosis di Jawa Barat secara luas, maka pemerintah dapat fokus dalam bidang lingkungan dan kesehatan. Hal ini dapat dengan melakukan pembangunan-pembangunan yang berfokus dalam memperbaiki kualitas lingkungan di Jawa Barat, dan fokus pula dalam fasilitas kesehatan secara keseluruhan di Provinsi Jawa Barat.

### 3.5 Ukuran Kebaikan Analisis LASSO dan *Group* LASSO

Hasil analisis LASSO dan *group* LASSO memiliki perbedaan. Salah satunya adalah banyak peubah yang digunakan dalam model LASSO jauh lebih sedikit

dibandingkan *group* LASSO. Hal ini akan berpengaruh pada nilai  $R_{adj}^2$  nya, dimana nilai  $R_{adj}^2$  dari *group* LASSO akan lebih kecil dibandingkan dengan LASSO. Ukuran kebaikan kedua model ini disajikan dalam Tabel 4.

Tabel 4: Ukuran kebaikan analisis LASSO dan *group* LASSO.

Ukuran	LASSO	<i>Group</i> LASSO
MSE	116.34	88.17
$R^2$	81.79%	85.27 %
$R_{adj}^2$	77.45%	68.09 %

MSE menunjukkan ukuran *error*, sehingga semakin kecil nilai MSE akan semakin baik. Penelitian ini menggunakan nilai MSE sebagai nilai pembanding utama. Nilai *R-squared* adalah seberapa besar keragaman dari peubah respon dapat dijelaskan oleh peubah penjelas, sehingga semakin besar ukuran ini, akan semakin baik. *Adjusted R-squared* merupakan nilai *R-squared* dengan mempertimbangkan banyak peubah penjelas yang masuk ke dalam model.

Terlihat dari Tabel 4, bahwa model *group* LASSO lebih baik di nilai MSE yang lebih rendah dan  $R^2$  yang lebih tinggi. Sedangkan, LASSO lebih baik pada  $R_{adj}^2$  yang lebih tinggi. Hal tersebut wajar terjadi, karena banyak peubah penjelas yang masuk dalam model *group* LASSO jauh lebih besar dibandingkan LASSO, sesuai dengan definisi  $R_{adj}^2$  yang memperhitungkan bias banyak parameter dalam model. Jika mengacu pada nilai ini, maka dapat dikatakan bahwa analisis *group* LASSO lebih baik. Selain itu, pada analisis *group* LASSO, dapat diketahui kelompok peubah mana yang paling berpengaruh dan tidak berpengaruh. Namun perbedaan yang tidak terlalu jauh ini dapat membuat peneliti dapat mengkolaborasikan kedua hasil analisis ini. Ketika yang difokuskan dan dibutuhkan adalah analisis yang melihat peubah penjelas yang dapat menjelaskan jumlah kasus tuberkulosis di Jawa Barat, maka dapat dilihat model pada LASSO. Hal ini karena model pada LASSO ini lebih mudah diinterpretasikan karena peubah penjelas yang masuk pada model lebih sedikit dibandingkan *group* LASSO. Ketika yang difokuskan dan dibutuhkan adalah analisis yang melihat kelompok peubah penjelas yang dapat menjelaskan jumlah kasus tuberkulosis di Jawa Barat, maka dapat dilihat model pada *group* LASSO. Hal ini karena model pada *group* LASSO ini adalah analisis yang lebih difokuskan pada analisis kelompok peubah penjelas.

### 3.6 Dari *Machine Learning* Menuju *Statistical Learning*

Pada analisis *machine learning*, pengujian hipotesis dan selang kepercayaan yang bersifat statistika tidak menjadi fokus. Hal ini membuat hasil dari analisis ini dipandang sebagai sesuatu yang deterministik. Padahal, dalam analisis secara statistik, selalu ada *error* yang harus diperhitungkan, sehingga tidak dapat langsung dikatakan bila suatu hasil prediksi tersebut bersifat mutlak. Hal ini menyebabkan dibutuhkannya analisis secara *statistical learning* dalam memerhitungkan nilai *error* yang mungkin muncul dalam suatu analisis. Metode LASSO (dan *group* LASSO) merupakan metode

dapat dikembangkan menjadi *statistical learning*, dengan melakukan *bootstrapping*. Pada proses ini, hasil model yang akan diuji secara *statistical learning* ini adalah model LASSO. Penggunaan model LASSO ini dilakukan karena lebih mudah, dibandingkan dengan model *group* LASSO. Hal ini karena jumlah peubah yang masuk ke dalam model LASSO lebih sedikit dibandingkan dengan model *group* LASSO. Ini dilakukan sebagai contoh bahwa analisis yang dikembangkan dalam *machine learning* seperti LASSO dan *group* LASSO, dapat diuji lebih lanjut menjadi *statistical learning*. Proses ini dilakukan untuk mendapatkan selang kepercayaan 95% untuk koefisien peubah penjelas. Selang kepercayaan ini digunakan untuk menguji hipotesis apakah nilai suatu koefisien suatu peubah penjelas yang dianalisis signifikan. Pada kasus ini, yang diuji adalah apakah koefisien peubah penjelas bernilai nol atau tidak. Jika koefisien bernilai nol, maka peubah penjelas tidak signifikan terhadap jumlah kasus tuberkulosis di Jawa Barat. Hipotesis yang diuji adalah:

$$H_0 : \beta_x = 0$$

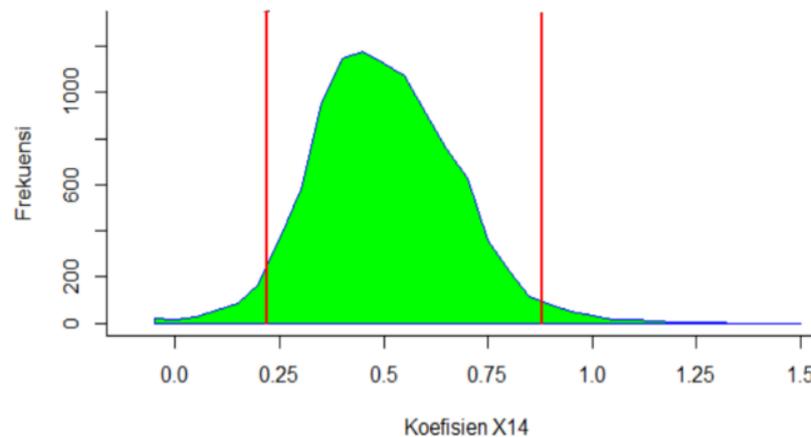
$$H_1 : \beta_x \neq 0$$

$H_0$  ditolak jika pada selang kepercayaan 95% ( $\alpha=5\%$ ) dari parameter  $\beta_x$  tidak mencakup nilai nol. Nilai  $\alpha$  merupakan peluang kesalahan dalam menolak  $H_0$ . Jika hasil pengujian hipotesis adalah menolak  $H_0$ , maka koefisien peubah penjelas ke- $x$  tersebut signifikan pada  $\alpha=5\%$ . Sebaliknya, jika hasil pengujian hipotesis adalah tidak menolak  $H_0$ , maka tidak dapat ditarik kesimpulan dikarenakan peluang kesalahan tidak menolak  $H_0$  tidak diketahui.

Penentuan selang kepercayaan 95% ini dilakukan dengan proses *bootstrapping*. Hal ini dilakukan dengan melakukan *resampling* sebanyak 27 sampel dari 27 kabupaten/kota dengan pengembalian. Sedangkan peubah penjelas yang diambil untuk masing-masing amatan adalah peubah penjelas hasil dari analisis LASSO sebelumnya, yaitu  $X_1$ ,  $X_5$ ,  $X_6$ ,  $X_{14}$  dan  $X_{23}$ . Satu set data tersebut kemudian dianalisis menggunakan analisis LASSO dengan lambda ( $\lambda$ ) yang ditentukan ulang melalui validasi silang *10-fold*, untuk mengetahui nilai koefisien dari masing-masing peubah. Hal ini diulang sebanyak 10 000 kali pengulangan. Ini untuk mendapatkan selang kepercayaan dan galat bakunya.

Penentuan selang kepercayaan 95% dapat dilakukan dengan melihat nilai batas bawah dan batas atasnya, yaitu data ke 2.5% dan 97.5% pada hasil 10 000 kali ulangan untuk suatu peubah penjelas. Pada  $X_1$ , selang kepercayaannya adalah  $0 \leq \beta_1 < 0.31$ . Pada  $X_5$ , selang kepercayaannya adalah  $0 \leq \beta_5 < 0.45$ . Pada  $X_6$ , selang kepercayaannya adalah  $-0.17 < \beta_6 < 0.39$ . Pada  $X_{14}$ , selang kepercayaannya adalah  $0.22 < \beta_{14} < 0.88$ . Terakhir adalah  $X_{23}$ , selang kepercayaannya adalah  $0 \leq \beta_{23} < 0.59$ .

Merujuk pada hasil ini, hasil pengujian hipotesis untuk  $X_1$ ,  $X_5$ ,  $X_6$  dan  $X_{23}$  pada  $\alpha=5\%$  adalah tidak menolak  $H_0$ . Ini berarti tidak ada kesimpulan untuk  $X_1$ ,  $X_5$ ,  $X_6$  dan  $X_{23}$  pada  $\alpha=5\%$ . Sedangkan hasil pengujian hipotesis untuk  $X_{14}$  adalah menolak  $H_0$ . Hal ini berarti peubah penjelas  $X_{14}$  signifikan pada  $\alpha=5\%$ .



Gambar 4: Kurva tebaran nilai koefisien  $X_{14}$  dari 10 000 ulangan melalui *bootstrap*.

Kurva tebaran koefisien  $X_{14}$  dapat dilihat pada Gambar 4. Garis merah pada kurva tebaran menunjukkan batas selang kepercayaan 95%. Selang 95% adalah selang di antara kedua garis merah pada kurva tebaran tersebut. Nilai simpangan baku dari koefisien  $X_1$ ,  $X_5$ ,  $X_6$ ,  $X_{14}$  dan  $X_{23}$  adalah 0.092, 0.123, 0.147, 0.17 dan 0.149. Sedangkan nilai ragam dari koefisien  $X_1$ ,  $X_5$ ,  $X_6$ ,  $X_{14}$  dan  $X_{23}$  adalah 0.008, 0.015, 0.021, 0.029 dan 0.022.

Berdasarkan hasil ini, diketahui bahwa melalui *bootstrap* dapat diketahui selang kepercayaan untuk masing-masing koefisien peubah penjelas dalam analisis LASSO. Hasil-hasil kurva tebaran secara eksplorasi juga menunjukkan koefisien untuk beberapa peubah penjelas, masih banyak terdapat nilai nol. Peubah dengan banyak koefisien yang bernilai nol seperti kurang kuat, karena peluang peubah tersebut tidak masuk ke dalam model masih tinggi. Pada  $\alpha=5\%$ , hanya peubah  $X_{14}$  yaitu jumlah orang bekerja yang keputusannya adalah signifikan. Hasil ini menunjukkan bahwa peubah  $X_{14}$  dapat menjelaskan jumlah kasus tuberkulosis di Jawa Barat menurut analisis LASSO.

#### 4. Simpulan

Analisis LASSO, dapat digunakan untuk menganalisis data dengan jumlah peubah penjelas yang banyak untuk menentukan peubah penjelas yang dapat menjelaskan jumlah kasus tuberkulosis di Jawa Barat. Sedangkan Analisis *group* LASSO dapat digunakan menentukan kelompok peubah yang dapat menjelaskan jumlah kasus tuberkulosis di Jawa Barat. Peubah yang dapat menjelaskan jumlah kasus tuberkulosis di Jawa Barat menurut analisis LASSO adalah jumlah orang difabel, jumlah tenaga kefarmasian, jumlah penderita gizi buruk, jumlah orang bekerja dan jumlah perkotaan. Sedangkan menurut *group* LASSO, kelompok peubah penjelas yang dapat menjelaskan jumlah kasus tuberkulosis di Jawa Barat adalah kelompok lingkungan dan kesehatan. Urutan kelompok peubah, mulai dari yang paling penting hingga tidak penting menurut analisis *group* LASSO adalah lingkungan, kesehatan, SDM dan yang terakhir adalah ekonomi. Pemerintah dapat menurunkan jumlah kasus

tuberkulosis di Jawa Barat dengan memperhatikan peubah dan kelompok peubah yang dapat menjelaskan jumlah kasus tuberkulosis di Jawa Barat.

## Daftar Pustaka

- Andini, W. (2018). 2018. Retrieved March 12, 2019, from Tuberkulosis (TBC), infeksi penyebab kematian nomor 1 di Indonesia. website: <http://hellosehat.com/pusat-kesehatan/tuberculosis-tbc/fakta-tbc-di-indonesia>.
- Dewi, Y. S. (2010). OLS, LASSO dan PLS pada data mengandung multikolinearitas. *Jurnal Ilmu Dasar*, 11(1): 83–91.
- Hastie, T., Tibshirani, R., & Wainwright, M. (2015). *Statistical learning with sparsity: the lasso and generalizations*. Florida (US): Chapman and Hall/CRC.
- Kinansi, R. R. (2017). Pendugaan data hilang pada rancangan percobaan uji daya bunuh ekstrak etanol akar tumbuhan tuba terhadap kecoa amerika (*Periplaneta americana*) menggunakan metode yates. *Buletin Penelitian Kesehatan*, 45(3): 205–214.
- Prabowo, F. K. H., Wilandari, Y., & Rusgiyono, A. (2015). Pemodelan pertumbuhan ekonomi jawa tengah menggunakan pendekatan least absolute shrinkage and selection operator (LASSO). *Jurnal Gaussian*, 4(4): 855–864.
- Pratama, W., & Wulandari, S. P. (2015). Pemetaan dan pemodelan jumlah kasus penyakit tuberkulosis (TBC) di provinsi Jawa Barat dengan pendekatan geographically weighted negative binomial regression (GWNBR). *Jurnal Sains Dan Seni ITS*, 4(1): D37–D42.
- Sejati, A., & Sofiana, L. (2015). Faktor-faktor terjadinya tuberkulosis. *KEMAS: Jurnal Kesehatan Masyarakat*, 10(2): 122–128.
- Soleh, Agus M, Wigena, A. H., Djuraidah, A., & Saefuddin, A. (2015). Statistical downscaling to predict monthly rainfall using linear regression with L1 regularization (LASSO). *Applied Mathematical Sciences*, 9(108): 5361–5369.
- Soleh, Agus Mohamad, & Aunuddin. (2013). LASSO: solusi alternatif seleksi peubah dan penyusutan koefisien model regresi linier. *Forum Statistika Dan Komputasi*, 18(1).
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1): 267–288.
- Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1): 49–67.