

PENERAPAN *GENERALIZED* LASSO DALAM PENDETEKSIAN PENCILAN PADA PEMODELAN HARGA RUMAH DI JAWA BARAT

SITI SHARA



DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2024

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Penerapan *Generalized LASSO* dalam Pendeteksian Pencilan pada Pemodelan Harga Rumah di Jawa Barat” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Agustus 2024

Siti Shara
G14180006

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



ABSTRAK

SITI SHARA. Penerapan *Generalized* LASSO dalam Pendeteksian Pencilan pada Pemodelan Harga Rumah di Jawa Barat. Dibimbing oleh MOHAMMAD MASJKUR dan SEPTIAN RAHARDIANTORO.

Pencilan adalah amatan yang secara signifikan berbeda jauh dari amatan lainnya dalam sebuah data. Keberadaan pencilan ini dapat mengganggu pendugaan koefisien dalam regresi. Pendeteksian pencilan bisa dilakukan dengan menggunakan metode *Generalized* Least Absolute Shrinkage Selection Operator (*Generalized* LASSO). *Generalized* LASSO adalah sebuah metode yang mengaplikasikan nilai D sebagai matriks penalti (D) untuk regresi lasso. Variasi dari nilai D beragam disesuaikan dengan kebutuhan penggunaan, salah satunya untuk mendeteksi pencilan. Pengaplikasian metode *Generalized* LASSO untuk mendeteksi pencilan pada pemodelan harga rumah di Jawa Barat tahun 2023 menunjukkan bahwa metode ini cukup baik dalam mendeteksi pencilan sehingga bisa digunakan sebagai alternatif untuk pendeteksian pencilan. Pemilihan lamda optimal berpengaruh terhadap jumlah amatan yang terdeteksi pencilan. Semakin besar nilai lamda yang digunakan maka model akan menjadi lebih sederhana dan amatan yang terdeteksi pencilan semakin sedikit. Sebaliknya, semakin kecil nilai lamda yang digunakan maka model akan menjadi semakin kompleks dan amatan yang terdeteksi pencilan lebih banyak.

Kata kunci: deteksi pencilan, *generalized* LASSO, pemodelan harga rumah

ABSTRACT

SITI SHARA. Application of *Generalized* LASSO in Outlier Detection in House Price Modelling in West Java. Supervised by MOHAMMAD MASJKUR and SEPTIAN RAHARDIANTORO.

Outliers are observations that are significantly different from other observations in the data. The existence of this outlier can interfere with the estimation of coefficients in regression. Outlier detection can be done using the *Generalized* Least Absolute Shrinkage Selection Operator (*Generalized* LASSO) method. *Generalized* LASSO is a method that applies the D value as a penalty matrix (D) for lasso regression. Variations in the D value vary according to usage needs, one of which is to detect outliers. The application of the *Generalized* LASSO method to detect outliers in house price modeling in West Java in 2023 shows that this method is quite good at detecting outliers so that it can be used as an alternative for outlier detection. Choosing the optimal lambda affects the number of observations that detect outliers. The greater the lambda value used, the simpler the model will be and the fewer outliers detected will be observed. On the other hand, the smaller the lambda value used, the more complex the model will become and more outliers will be detected.

Keywords: *generalized* LASSO, house price modelling, outlier detection

@HikmahPutriIPB University

IPB University

Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2024
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.

PENERAPAN *GENERALIZED* LASSO DALAM PENDETEKSIAN PENCILAN PADA PEMODELAN HARGA RUMAH DI JAWA BARAT

SITI SHARA

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana pada
Program Studi Statistika dan Sains Data

**DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2024**

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Tim Penguji pada Ujian Skripsi:
1 Prof. Dr. Ir. Anik Djuraidah, M.S.



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



Judul Skripsi : Penerapan *Generalized* LASSO dalam Pendeteksian Pencilan pada Pemodelan Harga Rumah di Jawa Barat

Nama : Siti Shara

NIM : G14180006

Disetujui oleh

Pembimbing 1:

Ir. Mohammad Masjkur, M.S

Pembimbing 2:

Dr. Septian Rahardianto, M. Si

Diketahui oleh

Ketua Departemen Statistika:

Dr. Bagus Sartono, M. Si

NIP. 19780411200501002

Tanggal Ujian:
28 Juni 2024

Tanggal Lulus:

Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



PRAKATA

Puji dan syukur penulis panjatkan kepada Allah subhanaahu wa ta'ala atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan dengan judul “Penerapan *Generalized LASSO* dalam Pendeteksian Pencilan pada Pemodelan Harga Rumah di Jawa Barat”.

Terima kasih penulis ucapkan kepada pihak-pihak yang telah memberikan bimbingan, saran, doa, dukungan, motivasi, dan bantuan kepada penulis dalam penyelesaian karya ilmiah ini.

1. Abi (Muh. Ikhwan) dan Umi (Juaeriah) selaku orang tua penulis, Kakak Agung, Kakak Netha, Adek Fadila, Adek Farid, Azwa, yang selalu memberikan doa, limpahan kasih sayang, nasihat, dan motivasi serta support moral dan materiil, kepada penulis dalam menyelesaikan karya ilmiah ini.
2. Bapak Ir. Mohammad Masjkur, M. S., dan Bapak Dr. Septian Rahardiantoro, M. Si., selaku dosen pembimbing yang telah senantiasa membimbing, memberikan saran dan masukan, dan memotivasi penulis untuk menyelesaikan karya ilmiah ini.
3. Ibu Prof. Dr. Ir. Anik Djuraidah, M.S. selaku penguji sidang skripsi yang telah memberikan saran dan masukan dalam karya ilmiah ini.
4. Seluruh dosen dan staff Tata Usaha Departemen Statistika IPB yang telah memberikan banyak ilmu dan pembelajaran yang bermanfaat selama perkuliahan serta menunjang kebutuhan administrasi penulis selama masa perkuliahan dan dalam penyelesaian karya ilmiah.
5. Yiyin, Farah, Ami, Zalfa, Cemara, dan teman-teman Statistika 55 lainnya atas kebersamaan, bantuan, dan dukungannya kepada penulis sedari awal masa perkuliahan hingga akhir.
6. Nunung, Luki, Nanik, Marlina, Diva, Kak Rini, Kak Omiq, Kak Malika dan teman-teman Himagis 55 lainnya yang selalu kebersamai, memberikan bantuan, dukungan dan motivasi serta menjadi keluarga kedua penulis selama di rantauan.
7. Kiki, Bang Bani, Arie, Santika, Tamimi, Iffa, Giri, dan Imam yang telah memberikan dukungan emosional, membantu, menemani, menghibur, dan memotivasi penulis dalam penyelesaian karya ilmiah ini.
8. Semua pihak yang tidak dapat penulis sebutkan satu-persatu atas segala kebaikan, doa, bantuan, dukungan, dan motivasi kepada penulis sehingga penulis dapat menyelesaikan karya ilmiah ini.

Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan.

Bogor, Agustus 2024

Siti Shara

DAFTAR ISI

DAFTAR TABEL	x
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	x
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan	2
1.3 Ruang Lingkup	2
II TINJAUAN PUSTAKA	3
2.1 Pendeteksian Pencilan	3
2.2 <i>Generalized</i> LASSO	3
2.3 Penentuan Tuning Parameter Optimum	4
2.4 Jarak Mahalanobis	4
2.5 Jarak Cook	5
III METODE	6
3.1 Bahan dan Data	6
3.2 Prosedur Analisis Data	6
IV HASIL DAN PEMBAHASAN	8
4.1 Standarisasi Data	8
4.2 Eksplorasi Data	8
4.3 Pendeteksian pencilan berdasarkan pendekatan multivariate	11
4.4 Pendeteksian Pencilan dengan <i>Generalized</i> LASSO	12
4.5 Perbandingan hasil deteksi pencilan menggunakan metode <i>Generalized</i> LASSO dengan jarak Mahalanobis dan jarak Cook	13
V SIMPULAN DAN SARAN	15
5.1 Simpulan	15
5.2 Saran	15
DAFTAR PUSTAKA	16
LAMPIRAN	18
RIWAYAT HIDUP	29



DAFTAR TABEL

1	Peubah penelitian	6
2	Perbandingan data sebelum dan sesudah distandarisasi	8
3	Rekapitulasi amatan yang terdeteksi pencilan pada boxplot	11
4	Korelasi antar peubah respon dengan peubah penjelas	11
5	Rangkuman amatan yang terdeteksi sebagai pencilan	12
6	Perbandingan hasil deteksi pencilan dengan berbagai metode	13

DAFTAR GAMBAR

1	Histogram masing-masing peubah yang digunakan dalam penelitian	9
2	Boxplot masing-masing peubah yang telah distandarisasi dan digunakan dalam penelitian ini	10

DAFTAR LAMPIRAN

1	Lampiran 1 Hasil vector koefisien $\tilde{\beta} = (\alpha, \beta)^T$ dengan $\lambda=2383,003$	19
2	Lampiran 2 Hasil vector koefisien $\tilde{\beta} = (\alpha, \beta)^T$ dengan $\lambda=1524,5$	21
3	Lampiran 3 Hasil vector koefisien $\tilde{\beta} = (\alpha, \beta)^T$ dengan $\lambda=1440$	23
4	Lampiran 4 Hasil vector koefisien $\tilde{\beta} = (\alpha, \beta)^T$ dengan $df=13$	25
5	Lampiran 5 Hasil vector koefisien $\tilde{\beta} = (\alpha, \beta)^T$ dengan $df=22$	27

Hak Cipta Dilindungi Undang-undang
 1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.