

METODE SPARSE GRUP LASSO UNTUK MEMODELKAN KECENDERUNGAN PERILAKU SEKSUAL REMAJA PUTRI DI PROVINSI BENGKULU

Hengki Muradi¹⁾, Khairil Anwar Notodiputro²⁾, Bagus Sartono³⁾

¹⁾Departemen Matematika Institut Sains Dan Teknologi Nasional, Jl. Moh Kahfi II,
Jakarta Selatan; hengki.muradi@istn.ac.id

²⁾Departemen Statistika Institut Pertanian Bogor, Jl. Raya Dramaga,
Bogor; khairil@apps.ipb.ac.id

³⁾Departemen Statistika Institut Pertanian Bogor, Jl. Raya Dramaga, Bogor;
bagusco@gmail.com

Abstract

The combination of the lasso variable selection method and the lasso group results in a variable selection method that can overcome problems in both the lasso method and the lasso group. This combination can also be applied to the cases of non-orthogonal groups as required in the lasso group method. This paper discusses the lasso sparse method with the coordinate descent approach as proposed by Friedman, et.al (2010). The real case raised in this study is the tendency of the sexual behaviour of young women in Bengkulu province. Because there are 15 explanatory variables for X , a variable selection process is needed to obtain the simplest model. The results were selected as many as 3 explanatory variables so that the remaining 12 explanatory variables can then be modelled in logistic regression. The substance of the unselected regression parameters is young women with more high school education, living in rural areas, rarely reading newspapers/magazines, rarely accessing the internet > 1 time a month, and strongly agree with virginity more likely to engage in sexual activity compared with educational characteristics, residence, newspaper reading intensity, internet access, and other virginity attitudes.

Keywords. Lasso, lasso groups, Sparse grup lasso, Sexual behavior, Models

1. Pendahuluan

Metode grup lasso merupakan perluasan dari metode lasso yang bertujuan untuk menyeleksi sekumpulan peubah penjelas dalam model, khususnya pada kasus banyaknya penjelas jauh lebih besar daripada banyaknya pengamatan ($p \gg n$). Grup lasso dimulai dengan membuat kelompok-kelompok dari sekumpulan peubah, kemudian diseleksi. Yuan & Lin (2007) berhasil memformulasikan algoritma metode grup lasso yang cara kerjanya

sama dengan metode lasso pada level grup. Pada level grup, optimasi bergantung pada parameter tuning λ . Jika $\lambda=0$ maka seluruh parameter dalam grup dieliminasi.

Menurut Friedman, et.al (2010) untuk metode grup lasso yang tidak mengalami sparsitas, jika parameter λ pada level grup bernilai nol maka seluruh parameter dalam grup akan tereliminasi. Karena itu diusulkan penambahan penalty pada grup lasso sehingga terdapat sparsitas dalam grup. Metode ini disebut dengan metode Sparse Group Lasso (SGL).

Metode SGL merupakan kombinasi dari lasso dan grup lasso. Friedman, et.al (2010) menyatakan bahwa algoritma SGL mampu menangani permasalahan sparsitas dan dapat bekerja pada kondisi matriks kovariat tidak ortonormal. Untuk mendukung pernyataan tersebut, Friedman, et.al (2010) membuat simulasi dengan data bangkitan sebanyak 200 observasi dan 100 peubah penjelas. Hasilnya menunjukkan bahwa metode SGL lebih efektif dibandingkan metode lasso dan grup lasso. Sayangnya, hasil implementasi data simulasi tidak selalu sejalan dengan kasus nyata. Oleh karena itu, dalam penelitian ini akan diimplementasikan metode sparse grup lasso untuk kasus pemodelan kecenderungan perilaku seksual pada remaja putri.

2. Metode Penelitian

2.1 Data Penelitian

Data penelitian yang digunakan pada penelitian ini adalah data SDKI tahun 2017 di provinsi Bengkulu. Peubah respon adalah pengakuan responden remaja putri yang berusia 15-24 tahun tentang bagaimana mereka berperilaku pada saat berinteraksi berdua dengan pasangannya sebelum menikah. Kovariat yang dimasukkan ke dalam proses peubah adalah pendidikan tertinggi, daerah tempat tinggal, intensitas membaca surat kabar/majalah, intensitas mengakses internet, alkohol, dan pandangan terhadap keperawanan. Secara rinci peubah-peubah tersebut disajikan pada Tabel 1 berikut ini.

Tabel 1. Peubah yang digunakan dalam pemodelan

Peubah	Notasi	Kategori	Skala Ukur
Perilaku Seksual Remaja Putri	y	0 = Berpegangan tangan atau berpelukan 1 = Berciuman atau meraba	Nominal
Pendidikan Tertinggi	X_1	1 = SD 2 = SMP 3 = SMA 4 = PT	Ordinal
Daerah Tempat Tinggal	X_2	0 = Perdesaan 1 = Perkotaan	Nominal
Intensitas Membaca Surat Kabar/Majalah	X_3	1 = Paling Sedikit Seminggu Sekali 2 = Jarang 3. Tidak Pernah	Nominal
Intensitas Akses Internet	X_4	1= Hampir Setiap Hari 2 = Paling Sedikit Seminggu sekali 3 = Paling Sedikit Sekali Sebulan 4 = Tidak sama sekali	Nominal
Sikap Terhadap Keperawanan	X_5	0 = Sangat Setuju 1 = Setuju 2 = Tidak Setuju	Ordinal

2.2 Model Regresi Logistik

Berdasarkan sebaran peubah respon yakni binomial, maka model yang akan dibangun adalah model regresi logistik, yaitu:

$$\log \left(\frac{\pi}{1-\pi} \right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 \quad (3)$$

Dimana π merupakan probabilitas kejadian sukses dari peubah respon (McCullagh & Nelder, 1983). Peubah X_1 adalah pendidikan tertinggi remaja putri, X_2 adalah daerah tempat tinggal, X_3 merupakan intensitas membaca surat kabar/majalah, X_4 adalah intensitas mengakses internet, dan X_5 adalah sikap terhadap keperawanan.

Seluruh peubah prediktor merupakan kategorik sehingga dalam analisis data didefinisikan peubah dummy yang berjumlah sebanyak 15 peubah dummy. Peubah pendidikan (X_1) didumikan menjadi empat peubah dummy, peubah X_2 menjadi satu peubah dummy, peubah X_3 didummikan menjadi 3 variabel dummy, peubah X_4 dibuat menjadi tiga peubah dummy, dan peubah X_5 menjadi tiga peubah dummy.

2.3 Tahapan Analisis Data

1. Eksplorasi data yang berkaitan dengan model penelitian
2. Cleaning data dari *missing value*.
3. Input data ke aplikasi R.
4. Paket di R yang digunakan untuk analisis data adalah paket SGL dengan cara kerja sebagai berikut:

a. Nilai awal : $\hat{\beta} = \beta_0$

Pada grup l didefinisikan residual

$$r_l = y - \sum_{k \neq l} X_k \beta_k$$

$$r_j = y' - \sum_{k \neq j} Z_k \theta_k$$

a) Jika $J(\hat{t}) = \sum_{j=1}^k s_j^2 \leq 1$ maka $\hat{\beta} = \beta_0$.

b) Jika $J(\hat{t}) > 1$ maka

- Jika $|Z_j^T r_j| < \lambda_2$ maka $\hat{\theta}_j = 0$
- Jika $|Z_j^T r_j| \geq \lambda_2$ maka minimalisasi fungsi:

$$\frac{1}{2} \sum_{i=1}^N \left(y_i' - \sum_{j=1}^k Z_{ij} \theta_j \right)^2 + \lambda_1 \|\theta\|_2 + \lambda_2 \sum_{j=1}^k |\theta_j|$$

6. Ulangi (2) sampai grup-grup l konvergen

3. Hasil dan Pembahasan

Secara deskriptif perilaku seksual dan kovariatnya dapat dilihat pada tabel 2 berikut ini:

Tabel 2. Gambaran Perilaku seksual remaja Putri dan Kovariatnya

Peubah	Kategori	Frekuensi	%
Perilaku Seksual Remaja Putri	Tidak pacaran, Berpegangan Tangan atau Berpelukan	114	90,48
	Berciuman atau meraba	12	9,52
Pendidikan Tertinggi	SD	3	2.38%
	SMP	15	11.90%
	SMA	60	47.62%
	PT	48	38.10%
Daerah Tempat Tinggal	Perdesaan	59	46.83%
	Perkotaan	67	53.17%
Intensitas Membaca Surat Kabar/Majalah	Paling Sedikit Seminggu Sekali	39	30.95%
	Jarang	59	46.83%
	Tidak pernah	28	22.22%
Intensitas Akses Internet	Hampir Setiap Hari	84	75.7%
	> 1 kali seminggu	20	18.0%
	> 1 kali sebulan	7	6.3%
Sikap Terhadap Keperawanan	Sangat Setuju	91	72.22%
	Setuju	34	26.98%
	Tidak Setuju	1	0.79%

Berdasarkan perilaku seksual diperoleh data bahwa dari 126 remaja putri di Provinsi Bengkulu yang di survei pada tahun 2017, sebanyak 114 responden (90,48%) mengakui hanya melakukan kegiatan berpegangan tangan atau berpelukan sedangkan 12 responden (9,52%) lainnya melakukan kegiatan berciuman dan meraba. Kegiatan berpegangan tangan atau berpelukan termasuk perilaku seksual ringan sedangkan kegiatan berciuman dan meraba merupakan perilaku seksual berat.

Pada proses seleksi peubah menggunakan metode sparse lasso, digunakan paket SGL yang dibangun oleh Simon.et.al (2018). Paket SGL tersebut dapat mengakomodasi model linear, logistic, dan Cox. Hasil seleksi peubah menggunakan metode seleksi sparse grup lasso adalah sebagai berikut:

Tabel 3. Hasil Seleksi Peubah Dengan Metode Sparse Grup Lasso

Peubah	Dummy Peubah	Parameter β	
		Initial Value	Hasil Akhir
Pendidikan Tertinggi	SD	0	-6,129
	SMP	0	-0,298
	SMA	0	4,790
	PT	0	0,000
Daerah Tempat Tinggal	Perdesaan	0	1,956
	Perkotaan	0	-1,398
Intensitas Membaca Surat Kabar/Majalah	Paling Sedikit Seminggu Sekali	0	0,000
	Jarang	0	2,224
	Tidak pernah	0	-7,660
Intensitas Akses Internet	Hampir Setiap Hari	0	-3,584
	> 1 kali seminggu	0	-6,161
	> 1 kali sebulan	0	6,328
Sikap Terhadap Keperawanan	Sangat Setuju	0	0,879
	Setuju	0	-0,587
	Tidak Setuju	0	0,000

Pada Tabel 3 dapat dilihat bahwa dari 15 peubah dummy yang diseleksi, diperoleh hasil bahwa dengan cara mengelompokkan peubah penjelas menjadi kelompok-kelompok yang berangotakan tiga peubah, terdapat tiga peubah dummy yang terseleksi yaitu pendidikan perguruan tinggi, intensitas membaca surat kabar/majalah paling sedikit seminggu sekali, dan peubah dummy sikap tidak setuju terhadap keperawanan. Jumlah iterasi yang di setting secara default oleh R adalah sebanyak 1000 iterasi yang menghasilkan parameter lambda (λ) sebesar 0,00022.

Dalam penelitian ini dilakukan uji coba dengan mengatur jumlah anggota masing-masing kelompok dengan berbagai macam variasi, yakni 1 anggota, 2 anggota, 4 anggota, dan 5 anggota. Hasilnya diperoleh bahwa walaupun jumlah anggota kelompok berbeda-beda namun hasilnya tetap konsisten.

Hasil penelitian ini menunjukkan bahwa metode sparse grup lasso dapat diimplementasikan dalam prosedur seleksi peubah. Menurut Fonti (2017), prosedur seleksi peubah sangat penting dan menantang dalam pemodelan, terutama karena output yang diinginkan bervariasi untuk set data yang berbeda, dan sulit untuk menemukan model yang bekerja untuk setiap jenis masalah. Tugas menjadi lebih menantang ketika berhadapan dengan dataset dimensi tinggi.

Berdasarkan koefisien model regresi logistik yang diperoleh dari proses seleksi sparse grup lasso dapat dijelaskan bahwa berdasarkan pendidikan, remaja putri yang berpendidikan SMA lebih berpeluang melakukan aktivitas seksual berupa berciuman atau meraba dibandingkan dengan remaja putri yang berpendidikan SD maupun SMP. Berdasarkan daerah tempat tinggal, remaja putri yang tinggal di perdesaan lebih berpeluang melakukan aktivitas seksual dibandingkan dengan remaja putri yang tinggal di perkotaan. Remaja putri yang jarang membaca lebih berpeluang melakukan aktivitas seksual dan remaja putri yang mengakses internet > 1 kali dalam satu bulan lebih berpeluang dibandingkan dengan remaja putri yang lebih intens dalam mengakses internet. Dan remaja putri yang sangat setuju dengan keperawanan justru lebih berpeluang melakukan aktivitas seksual dibandingkan dengan remaja putri yang setuju dengan keperawanan.

4. Diskusi

Dalam penelitian ini, parameter tuning (λ) di set $\lambda_1 = \lambda_2$. Akan menjadi lebih menarik jika parameter tuning ini di set tidak sama. Efek dari pengaturan tersebut menarik untuk diteliti, apakah akan memberikan hasil seleksi optimal yang sama.

Berkaitan dengan matriks kovariat, dalam penelitian ini masing-masing grup berisi vektor-vektor yang saling orthogonal. Friedman, et.al (2010) mengklaim bahwa algoritma sparse grup lasso yang di usulkannya mampu di implementasikan pada kasus dimana vektor-vektor yang tidak saling orthogonal. Karena itu, penelitian yang mengimplementasikan kasus dimana vektor-vektor yang tidak saling orthogonal dapat membuktikan klain tersebut.

5. Kesimpulan dan Saran

Metode seleksi sparse grup lasso dengan menggunakan pendekatan *coordinate descent* dapat diimplementasikan pada kasus regresi logistik. Dalam penelitian ini, dari 20 peubah yang di seleksi terdapat 3 peubah yang terseleksi dengan nilai parameter tuning $\lambda = 0,00022$. Ketiga variabel yang terseleksi adalah pendidikan PT, intensitas membaca surat kabar paling sedikit satu kali satu minggu, dan sikap tidak setuju terhadap keperawanan.

Penelitian ini fokus kepada implementasi metode seleksi sparse grup lasso dengan asumsi bahwa matriks kovariat bersifat orthogonal. Friedman, et.al (2010) mengklaim bahwa algoritma yang sparse grup lasso yang ia diusulkan mampu mengakomodasi pada model dengan asumsi bahwa matriks kovariat tidak orthogonal. Oleh karena disarankan kepada peneliti berikutnya untuk dapat memvalidasi klaim tersebut dengan melakukan implementasi pada kasus real.

Daftar Pustaka

- Bakin, S. 1999. *Adaptive regression and model selection in data mining problems*. PhD Thesis. Australian National University, Canberra.
- Jerome Friedman, Trevor Hastie, & Robert Tibshirani. 2010. *A note on the group lasso and a sparse group lasso*. arXiv: 1001.0736v1.

- McCullagh, P & J.A. 1983. Nelder FRS. *Generalized Linear Models; Second edition*. London: Chapman and Hall.
- Noah Simon, Jerome Friedman, Trevor Hastie, and Rob Tibshirani. 2018. *Package 'SGL'*.CRAN Documentation.
- Sarwono. 2007. *Psikologi Remaja*. Jakarta : Raja Grafindo Persada.
- Tjiptaningrum, R. 2009. *Psikologi Remaja dan Kesehatan Reproduksi*. Jakarta: Grafindo Persada.
- Valeria Fonti. 2017. *Feature Selection using LASSO*. Research Paper in Business Analytics VU Amsterdam.
- Yuan, M. & Lin, Y. 2007. *Model selection and estimation in regression with grouped variables*. Journal of the Royal Statistical Society, Series B 68 (1), 49–67.