

PEMODELAN BERBASIS WILAYAH DATA DIMENSI TINGGI PADA KASUS *STUNTING* DI INDONESIA MENGGUNAKAN *GENERALIZED LASSO* TERMODIFIKASI

AIDA DARAJATI



PROGRAM STUDI SARJANA STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2025



@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 “Pemodelan Berbasis Wilayah Data Dimensi Tinggi pada Kasus *Stunting* di Indonesia Menggunakan *Generalized LASSO* Termodifikasi” 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, Juli 2025

Aida Darajati
G1401211016

- 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

AIDA DARAJATI. Pemodelan Berbasis Wilayah Data Dimensi Tinggi pada Kasus *Stunting* di Indonesia Menggunakan *Generalized LASSO* Termodifikasi. Dibimbing oleh SEPTIAN RAHARDIANTORO dan HARI WIJAYANTO.

Stunting masih menjadi permasalahan kesehatan di Indonesia, meskipun prevalensinya menurun hingga 6,1% di tahun 2024, beberapa provinsi masih memiliki tingkat prevalensi *stunting* sangat tinggi. Penelitian ini bertujuan untuk mengeksplorasi pola *stunting* di seluruh wilayah Indonesia, menelaah performa model *generalized LASSO* dalam identifikasi dugaan koefisien wilayah yang berdekatan pada beberapa ukuran ketetanggaan, serta menjelaskan peubah-peubah yang paling berpengaruh terhadap *stunting* di Indonesia. Data bersumber dari Badan Pusat Statistik dan Kementerian Dalam Negeri tahun 2024, mencakup 38 provinsi dengan sepuluh peubah bebas. Analisis dilakukan menggunakan modifikasi *elastic net* pada *generalized LASSO* melalui penambahan komponen penalti khusus (D_2) ke dalam regularisasi L_2 untuk mengurangi dampak korelasi antar peubah bebas. Pemilihan parameter *tuning* optimal dilakukan menggunakan metode ALOCV. Model KNN ($k = 2$) dengan matriks D_2 terpilih sebagai model terbaik berdasarkan kombinasi optimal dari derajat bebas, *CV-error*, dan nilai RMSE. Hasil analisis menunjukkan bahwa peubah *gini ratio* (Papua), *unmet need* pelayanan kesehatan (Sulawesi, Maluku, NTT), akses sanitasi layak (Sumatra, Kalimantan, Jawa bagian barat), dan ASI eksklusif (Sumatra, Jawa bagian barat) merupakan faktor paling berpengaruh terhadap prevalensi *stunting* di Indonesia tahun 2024. Dalam penelitian ini, penambahan matriks D_2 yang diusulkan mampu mengecilkan *CV-error* melalui identifikasi dampak serupa dari pasangan peubah bebas berkorelasi tinggi terhadap *stunting*.

Kata Kunci: ALOCV, dimensi tinggi, *generalized LASSO*, KNN, *stunting*.

ABSTRACT

AIDA DARAJATI. Region-Based Modeling of High-Dimensional Data on Stunting Cases in Indonesia Using a Modified Generalized LASSO. Supervised by SEPTIAN RAHARDIANTORO and HARI WIJAYANTO.

Stunting remains a health concern in Indonesia despite its prevalence decreasing to 6.1% in 2024, several provinces still have a very high prevalence of stunting. This research aims to explore the spatial patterns of stunting across Indonesia, examine the performance of the generalized LASSO model in identifying estimated adjacent region coefficients on several neighborhood measures, and explain the most influential variables on stunting prevalence in Indonesia. The data used is sourced from the Statistics Indonesia (BPS) and the Ministry of Home Affairs in 2024, covering 38 provinces with ten independent variables. The analysis was conducted using an elastic net modification on generalized LASSO by adding a specific penalty component ($\mathbf{D}_2$) to the L_2 regularization in order to reduce the effect of correlation among independent variables. Optimal tuning parameters were selected using the ALOCV method. The K-Nearest Neighbors (KNN) model with $k = 2$ and the $\mathbf{D}_2$ matrix was selected as the best model based on an optimal combination of degrees of freedom, lowest CV-error, and RMSE value. The analysis identified the gini ratio (Papua), unmet need for healthcare services (Sulawesi, Maluku, East Nusa Tenggara), access to proper sanitation (Sumatra, Kalimantan, Jakarta, Banten, West Java), and exclusive breastfeeding (Sumatra, Jakarta, Banten, West Java) as the most influential variables on stunting prevalence in 2024. In this study, the addition of the proposed $\mathbf{D}_2$ matrix was able to reduce the CV-error by identifying similar effects from pairs of highly correlated predictor variables on stunting.

Keywords: ALOCV, high-dimensional data, generalized LASSO, KNN, stunting.



- 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, tahun 2025
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.

PEMODELAN BERBASIS WILAYAH DATA DIMENSI TINGGI PADA KASUS *STUNTING* DI INDONESIA MENGGUNAKAN *GENERALIZED LASSO* TERMODIFIKASI

AIDA DARAJATI

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

**PROGRAM STUDI SARJANA STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2025**



@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:
Dr. Farit Mochamad Afendi, S.Si., M.Si.

Judul Skripsi : Pemodelan Berbasis Wilayah Data Dimensi Tinggi pada Kasus
Stunting di Indonesia Menggunakan *Generalized LASSO*
Termodifikasi

Nama : Aida Darajati
NIM : G1401211016

Disetujui oleh

Pembimbing 1:
Dr. Septian Rahardiantoro, S.Stat., M.Si.



Pembimbing 2:
Prof. Dr. Ir. Hari Wijayanto, M.Si.



Diketahui oleh

Ketua Program Studi Statistika dan Sains Data:
Dr. Bagus Sartono, S.Si., M.Si.
NIP 19780411 200501 1002



Tanggal Ujian:
21 Juli 2025

Tanggal Lulus:



- 

PRAKATA

Puji dan syukur penulis panjatkan kepada Allah subhanaahu wa ta'ala atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan tepat waktu. Penelitian yang dilaksanakan sejak bulan Januari 2025 sampai bulan Juli 2025 ini berjudul “Pemodelan Berbasis Wilayah Data Dimensi Tinggi pada Kasus *Stunting* di Indonesia Menggunakan *Generalized LASSO* Termodifikasi”.

Terima kasih penulis ucapkan sebanyak-banyaknya kepada seluruh pihak yang berperan aktif dalam membantu penulis untuk menyusun karya ilmiah ini, diantaranya:

1. Bapak Dr. Septian Rahardianto, S.Stat., M.Si. dan Bapak Prof. Dr. Ir. Hari Wijayanto, M.Si. selaku dosen pembimbing yang dengan penuh kesabaran telah memberikan begitu banyak bimbingan, arahan, dukungan, serta saran yang membangun sejak awal perencanaan penelitian hingga tahap penulisan ini selesai.
2. Seluruh keluarga terutama Ibu Laili Shofina, Bapak Agus Handoko, serta keluarga besar yang selalu memberikan dukungan, doa dan kasih sayangnya sehingga penulis dapat menyelesaikan karya ilmiah ini.
3. Seluruh dosen dan staf Departemen Statistika SSMI IPB yang membantu penulis selama perkuliahan sehingga dapat menyelesaikan studi dan administrasi sampai kelulusan.
4. Teman-teman Statistika 58, terutama Angga, Kamilah, Anis, Nadila, Dinda, Dindana, Hanifa, Ubaid, dan seluruh pihak yang tidak dapat disebutkan satu per satu yang telah banyak membantu peneliti dengan memberikan dukungan, doa, dan kebersamaan selama masa perkuliahan.

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

Bogor, Juli 2025

Aida Darajati



DAFTAR ISI

DAFTAR TABEL	ix
DAFTAR GAMBAR	ix
DAFTAR LAMPIRAN	ix
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan	3
II TINJAUAN PUSTAKA	4
2.1 Definisi Peubah Penelitian	4
2.2 Regularisasi <i>Elastic Net</i>	6
2.3 <i>Generalized LASSO</i>	7
2.4 <i>Generalized LASSO</i> Termodifikasi	7
2.5 <i>Approximate Leave-One-Out Cross Validation (ALOCV)</i>	8
III METODE	9
3.1 Data	9
3.2 Tahapan Analisis	9
IV HASIL DAN PEMBAHASAN	13
4.1 Eksplorasi dan Pra-Proses Data	13
4.2 Pembentukan Matriks Penalti D_1 dan D_2	15
4.3 Pemodelan dengan <i>Generalized LASSO</i>	17
4.4 Interpretasi Hasil dari Model Terbaik	18
4.5 Pemilihan Peubah Paling Berpengaruh terhadap <i>Stunting</i>	23
V SIMPULAN DAN SARAN	25
5.1 Simpulan	25
5.2 Saran	25
DAFTAR PUSTAKA	26
LAMPIRAN	29
RIWAYAT HIDUP	32



DAFTAR TABEL

1	Daftar peubah penelitian	9
2	Perbandingan model ALOCV	17
3	Koefisien tertinggi setiap peubah	23

DAFTAR GAMBAR

1	Ilustrasi <i>queen contiguity</i> di Pulau Papua	10
2	Peta prevalensi <i>stunting</i> di Indonesia tahun 2024	13
3	Sebaran <i>stunting</i> dengan (a) <i>density plot</i> dan (b) <i>QQ-plot</i> sebelum transformasi, (c) <i>density plot</i> dan (d) <i>QQ-plot</i> setelah transformasi	13
4	Matriks korelasi pearson antar peubah bebas	14
5	Plot ketetanggaan <i>queen contiguity</i>	15
6	Plot ketetanggaan KNN dengan (a) $k = 2$ dan (b) $k = 3$	16
7	Peta pengaruh penduduk miskin (X_1) terhadap <i>stunting</i>	18
8	Peta pengaruh <i>gini ratio</i> (X_2) terhadap <i>stunting</i>	19
9	Peta pengaruh JKI (X_3) terhadap <i>stunting</i>	19
10	Peta pengaruh JKA (X_4) terhadap <i>stunting</i>	20
11	Peta pengaruh <i>unmet need</i> (X_5) terhadap <i>stunting</i>	20
12	Peta pengaruh ASI eksklusif (X_6) terhadap <i>stunting</i>	21
13	Peta pengaruh sanitasi layak (X_7) terhadap <i>stunting</i>	21
14	Peta pengaruh air minum layak (X_8) terhadap <i>stunting</i>	22
15	Peta pengaruh ibu merokok (X_9) terhadap <i>stunting</i>	22
16	Peta pengaruh lama sekolah (X_{10}) terhadap <i>stunting</i>	23

DAFTAR LAMPIRAN

1	Lampiran 1 Transformasi Box-Cox	30
2	Lampiran 2 Plot nilai <i>CV-error</i> terhadap λ pada setiap model ALOCV	30
3	Lampiran 3 Pendugaan koefisien β pada model terbaik	31