

# PEMODELAN INTENSITAS KETERPAPARAN BANJIR DI SUMATRA BERBASIS PENGGEROMBOLAN TERBATAS SPASIAL DENGAN REGRESI LASSO DAN *GROUP* LASSO

CINDY APRILIYANTI



PROGRAM STUDI SARJANA STATISTIKA DAN SAINS DATA  
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA  
INSTITUT PERTANIAN BOGOR  
BOGOR  
2026



## PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Pemodelan Intensitas Keterpaparan Banjir di Sumatra Berbasis Penggerombolan Terbatas Spasial dengan Regresi LASSO dan *Group* LASSO” 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.

Dalam penyusunan karya ini, penulis menggunakan bantuan kecerdasan buatan ChatGPT untuk membantu penulisan karya ilmiah ini. Setelah menggunakan alat/layanan tersebut, penulis meninjau dan menyunting konten sesuai kebutuhan serta bertanggung jawab penuh atas isi karya tugas akhir ini.

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

Bogor, Juli 2026

Cindy Apriliyanti  
G1401221010



## ABSTRAK

CINDY APRILIYANTI. Pemodelan Intensitas Keterpaparan Banjir di Sumatra Berbasis Penggerombolan Terbatas Spasial dengan Regresi LASSO dan *Group* LASSO. Dibimbing oleh ERFIANI dan LA ODE ABDUL RAHMAN.

Banjir merupakan bencana hidrometeorologi yang paling sering terjadi di Indonesia dan menimbulkan dampak yang signifikan, terutama di wilayah Sumatra. Kompleksitas faktor penyebab banjir serta adanya heterogenitas spasial antarwilayah menjadi tantangan dalam pemodelan intensitas keterpaparan banjir. Penelitian ini bertujuan mengelompokkan kabupaten/kota di Sumatra berdasarkan karakteristik fisik wilayah. Faktor-faktor yang memengaruhi intensitas keterpaparan banjir selanjutnya diidentifikasi menggunakan LASSO dan *Group* LASSO. Perbandingan kinerja dilakukan antara model LASSO dan *Group* LASSO yang dibangun dengan *cluster* maupun non-*cluster*. Data yang digunakan terdiri atas 150 kabupaten/kota di Sumatra tahun 2024 dengan 24 peubah penjelas. Pengelompokan wilayah dilakukan menggunakan ClustGeo dan menghasilkan empat *cluster* dengan karakteristik topografi dan klimatologi yang berbeda. Pemodelan LASSO dan *Group* LASSO diterapkan baik pada data keseluruhan maupun per *cluster* wilayah. Model berbasis *cluster* memberikan kinerja yang lebih baik dibandingkan model non-*cluster*. Model *Cluster*-LASSO menghasilkan kinerja terbaik dengan RMSE sebesar 0,114 dan nilai  $R^2$  sebesar 63,6%. Hasil ini menunjukkan bahwa pengelompokan wilayah mampu meningkatkan kemampuan model dalam menjelaskan variasi intensitas keterpaparan banjir di Sumatra. Pengelompokan tersebut juga mampu mengidentifikasi faktor-faktor yang berbeda pada setiap karakteristik wilayah.

Kata kunci: *Cluster-Group* LASSO, *Cluster*-LASSO, ClustGeo, ClustOfVar, pemodelan bencana banjir



## ABSTRACT

CINDY APRILIYANTI. Modeling Flood Exposure Intensity in Sumatra Based on Spatial Limited Clustering with LASSO and Group LASSO Regression. Supervised by ERFIANI and LA ODE ABDUL RAHMAN.

Flood is the most frequently occurring hydrometeorological disaster in Indonesia and has significant impacts, particularly in Sumatra. The complexity of flood-causing factors and the presence of spatial heterogeneity across regions pose challenges in modeling flood exposure intensity. This study aimed to classify regencies and municipalities in Sumatra based on their physical characteristics. The factors affecting flood exposure intensity were identified using Least Absolute Shrinkage and Selection Operator (LASSO) and Group LASSO. The performance of LASSO and Group LASSO models with and without regional clustering was compared. The data consisted of 150 regencies and municipalities in Sumatra in 2024 with 24 explanatory variables. Regional clustering was performed using ClustGeo and produced four clusters with distinct topographic and climatological characteristics. LASSO and Group LASSO models were applied to both the entire dataset and each regional cluster. The cluster-based models provide better performance than the non-cluster models. The Cluster-LASSO model achieved the best performance, with an RMSE of 0.114 and an  $R^2$  of 63.6%. These findings indicate that regional clustering improves the model's ability to explain the variation in flood exposure intensity across Sumatra. Regional clustering also enables the identification of different influencing factors for each regional characteristic.

**Keywords:** Cluster-Group LASSO, Cluster-LASSO, ClustGeo, ClustOfVar, flood disaster modelling



@Hak cipta milik IPB 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 mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2026<sup>1</sup>  
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 INTENSITAS KETERPAPARAN BANJIR DI SUMATRA BERBASIS PENGGEROMBOLAN TERBATAS SPASIAL DENGAN REGRESI LASSO DAN *GROUP* LASSO**

**CINDY APRILIYANTI**

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

**PROGRAM STUDI STATISTIKA DAN SAINS DATA  
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA  
INSTITUT PERTANIAN BOGOR  
BOGOR  
2026**



@Hak cipta milik IPB University

IPB University

Tim Penguji pada Ujian Skripsi:  
1 Sachnaz Desta Oktarina, S.Stat., M.Agr., Ph.D.



IPB University  
— Bogor Indonesia —

Perpustakaan 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 : **Pemodelan Intensitas Keterpaparan Banjir di Sumatra Berbasis Penggerombolan Terbatas Spasial dengan Regresi LASSO dan Group LASSO**

Nama : **Cindy Apriliyanti**  
NIM : **G1401221010**

@Hak cipta milik IPB University

Disetujui oleh

Pembimbing 1:  
**Dr. Ir. Erfiani, M.Si**



Pembimbing 2:  
**La Ode Abdul Rahman, S.Si., M.Si**



Diketahui oleh

Ketua Program Studi:  
**Dr. Bagus Sartono, M.Si**  
**197804112005011002**

---

Tanggal Ujian:  
**15 Juli 2026**

Tanggal Lulus:



## PRAKATA

Puji dan syukur penulis panjatkan kepada Allah subhanaahu wa ta'ala atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Penelitian ini dilaksanakan sejak bulan Januari 2025 hingga bulan Juli 2026 dengan judul “Pemodelan Intensitas Keterpaparan Banjir di Sumatra Berbasis Penggerombolan Terbatas Spasial dengan Regresi LASSO dan *Group LASSO*”

Terima kasih penulis ucapkan kepada seluruh pihak yang telah berperan dalam membantu penulis untuk menyusun karya ilmiah ini, di antaranya:

1. Allah SWT, Tuhan semesta alam yang telah memberikan pertolongan, rezeki berupa kesehatan, kelancaran, serta kemudahan sehingga penulis bisa menyelesaikan penelitian ini.
2. Ibu Dr. Ir. Erfiani, M.Si dan Bapak La Ode Abdul Rahman, S.Si., M.Si selaku komisi pembimbing yang telah memberikan saran dan arahan dalam penyelesaian karya ilmiah ini.
3. Bapak Dr. Bagus Sartono, M.Si, Ibu Dr. Yenni Angraini, S.Si., M.Si., Ibu Sachnaz Desta Oktarina, M.Agr, Ph.D, dan Bapak Dr. Septian Rahardiantoro, S.Stat, M.Si selaku dosen-dosen yang membantu dan memotivasi penulis untuk terus menulis karya ilmiah ini.
4. Seluruh keluarga atas doa dan dukungannya sehingga penulis dapat menyelesaikan karya ilmiah ini.
5. Seluruh dosen, staf, dan TU Prodi Statistika SSMI IPB yang membantu penulis selama perkuliahan sehingga dapat menyelesaikan pendidikan dan lancarnya administrasi hingga tahap akhir.
6. Delia Fitri Audina, saudari tak sedarah penulis yang selalu menjadi *comfort place* penulis terutama selama penulisan karya ilmiah ini.
7. Pemilik NIM G1401221112 yang selalu hadir pada setiap proses pembelajaran di kampus sebagai *main support system* penulis.
8. Semua sahabat penulis yang selalu memberikan dukungan serta apresiasi untuk semua langkah yang dilewati penulis.
9. *Youtuber* favorit penulis, yakni Windah Basudara, Nadia Omara, dan Nessie Judge yang telah membantu memberikan asupan dopamin sehingga penulis kuat dalam menulis karya ilmiah ini.
10. Teman-teman angkatan statistika 59 yang telah memberikan dukungan dan memberikan kebersamaan selama masa-masa kuliah dari semester awal hingga memasuki semester akhir.
11. Seluruh pihak yang tidak mungkin disebutkan satu per satu yang telah memberikan saran, arahan, dan dukungan kepada penulis sehingga dapat menyelesaikan tugas akhir.

Tidak lupa penulis memohon maaf atas segala kekurangan dalam penyusunan karya ilmiah ini. Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan.

Bogor, Juli 2026

Cindy Apriliyanti



## DAFTAR ISI

|                                                                          |    |
|--------------------------------------------------------------------------|----|
| DAFTAR TABEL                                                             | ix |
| DAFTAR GAMBAR                                                            | ix |
| DAFTAR LAMPIRAN                                                          | x  |
| PENDAHULUAN                                                              | 1  |
| 1.1 Latar Belakang                                                       | 1  |
| 1.2 Tujuan                                                               | 2  |
| TINJAUAN PUSTAKA                                                         | 3  |
| 2.1 Konsep Intensitas Keterpaparan Banjir                                | 3  |
| 2.2 Heterogenitas Spasial                                                | 3  |
| 2.3 <i>Package</i> ClustGeo                                              | 3  |
| 2.4 <i>Package</i> ClustOfVar                                            | 5  |
| 2.5 LASSO ( <i>Least Absolute Shrinkage and Selection Operator</i> )     | 6  |
| 2.6 <i>Group</i> LASSO                                                   | 7  |
| 2.7 <i>K-Fold Cross Validation</i>                                       | 8  |
| 2.8 <i>Ordinary</i> Kriging                                              | 8  |
| 2.9 Algoritma Estimasi Koefisien pada LASSO dan <i>Group</i> LASSO       | 9  |
| 2.10 Metode Pemilihan Jumlah <i>Cluster</i> Terbaik                      | 10 |
| 2.11 Sumber Pustaka Peubah Penjelas                                      | 11 |
| III METODE                                                               | 12 |
| 3.1 Data                                                                 | 12 |
| 3.2 Prosedur Analisis                                                    | 13 |
| IV HASIL DAN PEMBAHASAN                                                  | 16 |
| 4.1 Interpolasi Data Hilang                                              | 16 |
| 4.2 Eksplorasi Data                                                      | 21 |
| 4.3 <i>Clustering</i> Wilayah dengan ‘ClustGeo’                          | 23 |
| 4.4 Karakteristik <i>Cluster</i> Wilayah                                 | 24 |
| 4.5 <i>Grouping</i> Peubah Penjelas dengan ‘ClustOfVar’                  | 26 |
| 4.6 Pemodelan LASSO Non- <i>Cluster</i>                                  | 27 |
| 4.7 Pemodelan <i>Group</i> LASSO Non- <i>Cluster</i>                     | 28 |
| 4.8 Pemodelan <i>Cluster</i> -LASSO                                      | 30 |
| 4.9 Pemodelan <i>Cluster-Group</i> LASSO                                 | 31 |
| 4.10 Evaluasi Total <i>Cluster</i> -LASSO dan <i>Cluster-Group</i> LASSO | 32 |
| 4.11 Perbandingan Evaluasi Model                                         | 33 |
| 4.12 Interpretasi Faktor yang Memengaruhi Keterpaparan Banjir            | 33 |
| V SIMPULAN DAN SARAN                                                     | 39 |
| 5.1 Simpulan                                                             | 39 |
| 5.2 Saran                                                                | 39 |
| DAFTAR PUSTAKA                                                           | 41 |
| LAMPIRAN                                                                 | 45 |
| RIWAYAT HIDUP                                                            | 60 |



## DAFTAR TABEL

|    |                                                                              |    |
|----|------------------------------------------------------------------------------|----|
| 1  | Sumber pustaka peubah penjelas                                               | 11 |
| 2  | Peubah-peubah yang digunakan dalam <i>clustering</i>                         | 12 |
| 3  | Peubah yang digunakan pada LASSO dan <i>Group LASSO</i>                      | 12 |
| 4  | Hasil regresi peubah <i>X1</i>                                               | 16 |
| 5  | Hasil evaluasi semivariogram teoritis <i>X1</i>                              | 17 |
| 6  | Parameter semivariogram exponential <i>X1</i>                                | 18 |
| 7  | Hasil regresi peubah <i>X2</i>                                               | 18 |
| 8  | Hasil evaluasi semivariogram teoritis <i>X2</i>                              | 20 |
| 9  | Parameter semivariogram exponential <i>X2</i>                                | 20 |
| 10 | Statistik deskriptif peubah <i>clustering</i>                                | 21 |
| 11 | Rata-rata peubah atribut tiap <i>cluster</i>                                 | 24 |
| 12 | Klasifikasi morfologi dan kemiringan lereng (Van Zuidam, 1985)               | 24 |
| 13 | Rangkuman karakteristik <i>cluster</i> wilayah                               | 25 |
| 14 | Hasil <i>grouping</i> peubah penjelas                                        | 26 |
| 15 | Nilai penduga koefisien hasil regresi LASSO non- <i>cluster</i>              | 28 |
| 16 | Nilai penduga koefisien hasil regresi <i>Group LASSO</i> non- <i>cluster</i> | 29 |
| 17 | Perbandingan jumlah observasi/data <i>test</i> pada proses <i>K-fold CV</i>  | 30 |
| 18 | Perbandingan MSE CV dan $\lambda$ optimum pada <i>Cluster-LASSO</i>          | 30 |
| 19 | Peubah terpilih dan evaluasi model <i>Cluster-LASSO</i>                      | 31 |
| 20 | Perbandingan MSE CV dan $\lambda$ optimum <i>Cluster-Group LASSO</i>         | 31 |
| 21 | Peubah terpilih dan evaluasi model <i>Cluster-Group LASSO</i>                | 32 |
| 22 | Evaluasi total <i>Cluster-LASSO</i> dan <i>Cluster-Group LASSO</i>           | 33 |
| 23 | Evaluasi model keseluruhan                                                   | 33 |
| 24 | Ringkasan interpretasi <i>Cluster 1</i>                                      | 34 |
| 25 | Ringkasan interpretasi <i>Cluster 2</i>                                      | 35 |
| 26 | Ringkasan interpretasi <i>Cluster 3</i>                                      | 36 |
| 27 | Ringkasan interpretasi <i>Cluster 4</i>                                      | 37 |

## DAFTAR GAMBAR

|    |                                                                        |    |
|----|------------------------------------------------------------------------|----|
| 1  | Diagram alir penelitian                                                | 15 |
| 2  | <i>Scatterplot</i> peubah <i>X1</i> menurut a) Bujur, b) Lintang       | 16 |
| 3  | Semivariogram eksperimental <i>X1</i>                                  | 17 |
| 4  | Semivariogram exponential <i>X1</i>                                    | 17 |
| 5  | <i>Scatterplot</i> peubah <i>X2</i> menurut a) Bujur, b) Lintang       | 18 |
| 6  | Semivariogram eksperimental <i>X2</i>                                  | 19 |
| 7  | Semivariogram exponential <i>X2</i>                                    | 20 |
| 8  | Histogram peubah respon                                                | 21 |
| 9  | <i>Heatmap</i> rata-rata elevasi                                       | 22 |
| 10 | Matriks korelasi peubah penjelas                                       | 22 |
| 11 | Jumlah <i>cluster</i> optimum (a) <i>Elbow</i> , (b) <i>Silhouette</i> | 23 |
| 12 | Grafik <i>trade-off pseudo-inertia</i> berdasarkan nilai $\alpha$      | 23 |

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.



## DAFTAR LAMPIRAN

|    |                                                                  |    |
|----|------------------------------------------------------------------|----|
| 1  | <i>Heatmap</i> hasil <i>clustering</i> wilayah dengan ClustGeo   | 46 |
| 2  | Dendogram hasil <i>grouping</i> ClustOfVar                       | 46 |
| 3  | Ringkasan interpretasi <i>Cluster</i> 1                          | 47 |
| 4  | Ringkasan interpretasi <i>Cluster</i> 2                          | 49 |
| 5  | Ringkasan interpretasi <i>Cluster</i> 3                          | 50 |
| 6  | Ringkasan interpretasi <i>Cluster</i> 4                          | 52 |
| 7  | Nilai penduga koefisien hasil regresi <i>Cluster</i> -LASSO      | 53 |
| 8  | Arah penduga koefisien <i>Cluster</i> -LASSO                     | 54 |
| 9  | Persamaan LASSO Non- <i>Cluster</i>                              | 55 |
| 10 | Persamaan <i>Cluster</i> -LASSO per <i>cluster</i>               | 55 |
| 11 | Nilai penduga koefisien <i>Cluster</i> - <i>Group</i> LASSO      | 56 |
| 12 | Arah penduga koefisien <i>Cluster</i> - <i>Group</i> LASSO       | 57 |
| 13 | Persamaan <i>Group</i> -LASSO Non- <i>Cluster</i>                | 59 |
| 14 | Persamaan <i>Cluster</i> - <i>Group</i> LASSO per <i>cluster</i> | 59 |