



SKRIPSI

KLASTERISASI WILAYAH TERDAMPAK BANJIR DI INDONESIA MENGGUNAKAN METODE *HIERARCHICAL DENSITY-BASED SPATIAL CLUSTERING OF APPLICATIONS WITH NOISE* (HDBSCAN) DAN *BAYESIAN OPTIMIZATION*

NASYWA AZZAH NABILA
NPM 22083010051

DOSEN PEMBIMBING
Aviolla Terza Damaliana, S.Si., M.Stat.
Shindi Shella May Wara, M.Stat.

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SAINS DATA
SURABAYA
2026**

LEMBAR PENGESAHAN

KLASTERISASI WILAYAH TERDAMPAK BANJIR DI INDONESIA MENGUNAKAN METODE *HIERARCHICAL DENSITY-BASED SPATIAL CLUSTERING OF APPLICATIONS WITH NOISE* (HDBSCAN) DAN *BAYESIAN OPTIMIZATION*

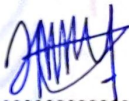
Oleh:
NASYWA AZZAH NABILA
NPM. 22083010051

Telah dipertahankan di hadapan dan diterima oleh Tim Penguji Sidang Skripsi Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur pada tanggal 10 Juni 2026:

Menyetujui,

**Aviolla Terza Damaliana, S.Si.,
M.Stat.**

NIP. 19940802 202203 2 015


.....

(Pembimbing I)

**Shindi Shella May Wara,
M.Stat.**

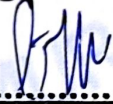
NIP. 19960518 202406 2 003


.....

(Pembimbing II)

Trimono, S.Si., M.Si.

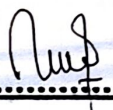
NIP. 19950908 202203 1 003


.....

(Ketua Penguji)

Muhammad Nasrudin, M.Stat.

NIP. 19960909 202406 1 002


.....

(Penguji I)

Mengetahui,

Dekan Fakultas Ilmu Komputer



Prof. Dr. Ir. Novirina Hendrasarie, MT.

NIP. 19681126 199403 2 001

LEMBAR PERSETUJUAN

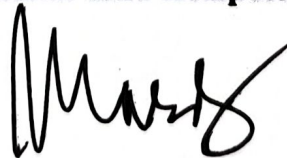
**KLASTERISASI WILAYAH TERDAMPAK BANJIR DI INDONESIA
MENGUNAKAN METODE *HIERARCHICAL DENSITY-BASED
SPATIAL CLUSTERING OF APPLICATIONS WITH NOISE (HDBSCAN)*
DAN *BAYESIAN OPTIMIZATION***

Oleh:
NASYWA AZZAH NABILA
NPM. 22083010051

Telah disetujui untuk mengikuti Ujian Skripsi

Menyetujui,

**Plt Koordinator Program Studi Sains Data
Fakultas Ilmu Komputer**



Dr. I Gede Susrama Mas Diyasa, ST., MT.
NIP. 19700619 2021211 009

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan di bawah ini:

Nama Mahasiswa : Nasywa Azzah Nabila
NPM : 22083010051
Program : Sarjana (S1)
Program Studi : Sains Data
Fakultas : Fakultas Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan secara lengkap dalam daftar pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila di kemudian hari ditemukan indikasi plagiat pada Skripsi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya.



Surabaya, 30 Juni 2026
Yang Membuat Pernyataan,



NASYWA AZZAH NABILA
NPM. 22083010051

ABSTRAK

Nama Mahasiswa / NPM : Nasywa Azzah Nabila / 22083010051
Judul Skripsi : Klasterisasi Wilayah Terdampak Banjir di Indonesia Menggunakan Metode *Hierarchical Density-Based Spatial Clustering of Applications with Noise* (HDBSCAN) dan *Bayesian Optimization*
Dosen Pembimbing : 1. Aviolla Terza Damaliana, S.Si., M.Stat.
2. Shindi Shella May Wara, M.Stat.

Banjir merupakan bencana alam yang paling sering terjadi di Indonesia. Berdasarkan data BNPB, tercatat 1.255 kejadian banjir pada tahun 2023 dan meningkat menjadi 1.420 kejadian pada tahun 2024. Peningkatan jumlah kejadian tersebut diikuti oleh kenaikan jumlah korban dari 3.860.136 jiwa menjadi 6.391.056 jiwa serta kenaikan jumlah kerusakan dari 1.083.004 unit menjadi 2.005.567 unit. Selain itu, setiap provinsi memiliki karakteristik dampak banjir yang berbeda-beda, baik dari segi korban jiwa maupun kerusakan infrastruktur yang ditimbulkan. Apabila perbedaan karakteristik dampak ini tidak dianalisis secara mendalam, maka alokasi sumber daya dalam upaya mitigasi bencana banjir berpotensi menjadi kurang efektif dan tidak tepat sasaran. Oleh karena itu, penelitian ini bertujuan untuk mengelompokkan provinsi di Indonesia berdasarkan indikator dampak banjir yang meliputi jumlah korban, kerusakan rumah, dan rumah terendam. Fokus penelitian ini adalah menerapkan metode *Hierarchical Density-Based Spatial Clustering of Applications with Noise* (HDBSCAN) yang dioptimasi menggunakan *Bayesian Optimization* untuk memperoleh kombinasi parameter optimal melalui pendekatan model probabilistik. Metode HDBSCAN dipilih karena mampu mengelompokkan data dengan kepadatan bervariasi sekaligus mengidentifikasi *noise* secara otomatis. Dengan parameter optimal $min_samples = 2$ dan $min_cluster_size = 2$, diperoleh nilai DBCV sebesar 0,5150 dan DCSI sebesar 0,7919 yang menghasilkan tiga kluster utama dan satu kluster *noise*. Kluster 0 berjumlah 5 provinsi, kluster 1 berjumlah 4 provinsi, kluster 2 berjumlah 21 provinsi, dan kluster *noise* berjumlah 8 provinsi. Penelitian ini juga mengembangkan GUI berbasis *website* interaktif dengan menggunakan *Streamlit* untuk memvisualisasikan hasil klasterisasi dalam bentuk peta interaktif.

Kata Kunci: Banjir, Klasterisasi, *Hierarchical Density-Based Spatial Clustering of Applications with Noise* (HDBSCAN), *Bayesian Optimization*, *Streamlit*

ABSTRACT

Student Name / NPM : Nasywa Azzah Nabila / 22083010051
Undergraduate thesis title : *Clustering of Flood-Affected Regions in Indonesia Using the Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) Method and Bayesian Optimization*
Advisors : 1. Aviolla Terza Damaliana, S.Si., M.Stat.
2. Shindi Shella May Wara, M.Stat.

Floods are the most frequently occurring natural disaster in Indonesia. According to data from BNPB, 1.255 flood events were recorded in 2023, increasing to 1.420 events in 2024. This increase in the number of events was followed by a rise in the number of affected people from 3.860.136 to 6.391.056, as well as an increase in the number of damaged units from 1.083.004 to 2.005.567. Furthermore, each province exhibits distinct flood impact characteristics, both in terms of human casualties and the resulting infrastructure damage. If these distinct impact characteristics are not analyzed in depth, resource allocation for flood disaster mitigation efforts risks being suboptimal and mistargeted. Therefore, this study aims to cluster provinces in Indonesia based on flood impact indicators, including the number of victims, damaged houses, and flooded houses. The focus of this study is the application of the Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) method, optimized using Bayesian Optimization to obtain optimal parameter combinations through a probabilistic modeling approach. HDBSCAN is chosen because it is capable of clustering data with varying densities while simultaneously identifying noise automatically. Using the optimal parameters of $\text{min_samples} = 2$ and $\text{min_cluster_size} = 2$, the model achieves a DBCV score of 0,5150 and a DCSI score of 0,7919, resulting in three main clusters and one noise cluster. Cluster 0 consists of 5 provinces, cluster 1 consists of 4 provinces, cluster 2 consists of 21 provinces, and the noise cluster consists of 8 provinces. Furthermore, this study develops a web-based interactive GUI using Streamlit to visualize the clustering results through interactive mapping.

Keywords: *Flood, Clustering, Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN), Bayesian Optimization, Streamlit*

KATA PENGANTAR

Puji syukur kehadirat Allah SWT atas segala rahmat, hidayah dan karunia-Nya kepada penulis sehingga skripsi dengan judul **“Klasterisasi Wilayah Terdampak Banjir di Indonesia Menggunakan Metode *Hierarchical Density-Based Spatial Clustering of Applications with Noise* (HDBSCAN) dan *Bayesian Optimization*”** dapat terselesaikan dengan baik. Dalam penyusunan skripsi ini, penulis ingin menyampaikan rasa terima kasih kepada berbagai pihak yang telah memberikan bantuan dan dukungan. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Ibu Prof. Dr. Ir. Novirina Hendrasarie, M.T., selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
2. Bapak Dr. Eng. Ir. Dwi Arman Prasetya, S.T., M.T., IPU., Asean Eng. selaku Ketua Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Ibu Aviolla Terza Damaliana, S.Si., M.Stat. selaku Dosen Pembimbing 1 yang telah meluangkan waktunya untuk memberikan bimbingan, arahan, serta saran kepada penulis selama proses penyusunan skripsi ini.
4. Ibu Shindi Shella May Wara, M.Stat. selaku Dosen Pembimbing 2 yang telah meluangkan waktunya untuk memberikan bimbingan, bantuan, serta motivasi kepada penulis selama proses penyusunan skripsi ini.
5. Dosen-dosen Program Studi Sains Data yang telah membimbing penulis dari awal perkuliahan hingga saat ini.
6. Orang tua penulis yang senantiasa memberikan doa dan dukungan materiil, serta adik-adik penulis yang selalu memberikan doa, dukungan, dan hiburan sehingga penulis dapat tetap termotivasi dan bersemangat dalam menyelesaikan skripsi ini hingga akhir.
7. Teman masa kecil penulis, Valeria, yang telah menjadi teman tumbuh bersama di segala kondisi dan selalu hadir dalam memberikan dukungan, bantuan, dan semangat kepada penulis selama menempuh pendidikan dasar hingga penyelesaian skripsi ini.

8. Teman-teman SMA penulis, yaitu Nabila, Dhista, Rachel, dan Diva yang selalu mendengarkan dan menerima segala keluhan penulis di segala situasi apa pun, serta selalu memberikan dukungan dan semangat kepada penulis selama menempuh pendidikan di bangku perkuliahan ini.
9. Teman-teman bermain penulis, yaitu Aca, Bella, Anggi, dan Netta yang selalu memberikan hiburan, semangat, dan dukungan sehingga penulis dapat menjalani proses penyusunan skripsi dengan lebih baik.
10. Teman-teman penulis di perkuliahan, yaitu Talitha, Dinda, Fiya, Fajria, Adel, Nova, Savvy, Rani, Kanza, dan Lauren yang selalu memberikan bantuan dan motivasi, baik dalam belajar bersama maupun saling memberi dukungan selama menempuh pendidikan di bangku perkuliahan ini.
11. Rekan-rekan mahasiswa Program Studi Sains Data, terima kasih atas kerja sama dan dukungan yang telah diberikan selama menempuh pendidikan di perkuliahan hingga keberhasilan penyusunan skripsi ini.
12. Terakhir, diri penulis sendiri yang telah berjuang dan bertahan semaksimal mungkin selama proses perkuliahan hingga penyusunan skripsi ini. Terima kasih karena tidak pernah menyerah dalam menghadapi setiap proses dan berusaha menjadi pribadi yang lebih baik selama menempuh pendidikan hingga penyelesaian skripsi ini.

Penulis menyadari bahwa di dalam penyusunan skripsi ini banyak terdapat kekurangan. Untuk itu, kritik dan saran yang membangun dari semua pihak sangat diharapkan demi kesempurnaan penulisan skripsi ini. Akhirnya, dengan segala keterbatasan yang penulis miliki, semoga laporan ini dapat bermanfaat bagi semua pihak pada umumnya dan penulis pada khususnya.

Surabaya, Juni 2026

Penulis

DAFTAR ISI

HALAMAN SAMPUL.....	i
LEMBAR PENGESAHAN	ii
LEMBAR PERSETUJUAN	iii
SURAT PERNYATAAN BEBAS PLAGIASI.....	iv
ABSTRAK	v
<i>ABSTRACT</i>	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xii
DAFTAR TABEL	xiii
DAFTAR LAMPIRAN	xiv
DAFTAR NOTASI.....	xv
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	4
1.3. Batasan Masalah.....	5
1.4. Tujuan Penelitian	5
1.5. Manfaat Penelitian	6
BAB II TINJAUAN PUSTAKA.....	7
2.1. Penelitian Terdahulu	7
2.2. Kerangka Teori.....	13
2.2.1. Banjir.....	13
2.2.2. Standarisasi Data	15
2.2.3. Uji Kaiser-Meyer-Olkin	15
2.2.4. Uji Bartlett.....	16
2.2.5. Uji Multikolinieritas	17
2.2.6. <i>Local Outlier Factor</i>	17
2.2.7. <i>Principal Component Analysis</i>	19
2.2.8. Klasterisasi	20

2.2.9.	<i>Density-Based Spatial Clustering of Applications with Noise (DBSCAN)</i>	21
2.2.10.	<i>Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN)</i>	22
2.2.11.	<i>Bayesian Optimization</i>	25
2.2.12.	Evaluasi Model	30
2.2.13.	<i>Streamlit</i>	33
BAB III METODOLOGI PENELITIAN		35
3.1.	Variabel Penelitian dan Sumber Data	35
3.2.	Langkah Analisis	36
3.1.1.	Pengumpulan Data	37
3.1.2.	<i>Exploratory Data Analysis</i>	38
3.1.3.	Pra-Pemrosesan Data	38
3.2.3.	Pemodelan Klasterisasi	39
3.2.4.	Pembuatan GUI	40
3.3.	Desain Sistem	41
BAB IV HASIL DAN PEMBAHASAN		45
4.1.	Pengumpulan Data	45
4.2.	<i>Exploratory Data Analysis</i>	46
4.3.	Pra-Pemrosesan Data	47
4.3.1.	Pemeriksaan Nilai yang Hilang	47
4.3.2.	Standarisasi Data	48
4.3.3.	Uji Statistik	50
4.3.4.	Deteksi <i>Outlier</i>	53
4.3.5.	Reduksi Dimensi Variabel	55
4.4.	Pemodelan Klasterisasi HDBSCAN	59
4.5.	Pemodelan Klasterisasi HDBSCAN dan <i>Bayesian Optimization</i>	61
4.5.1.	Implementasi Perhitungan Manual <i>Bayesian Optimization</i>	61
4.5.2.	Pencarian Parameter Optimal dengan <i>Bayesian Optimization</i>	66
4.5.3.	Implementasi Perhitungan Manual HDBSCAN	72
4.5.4.	Melakukan Klasterisasi HDBSCAN	78

4.5.5.	Evaluasi Model Klasterisasi.....	80
4.6.	Perbandingan HDBSCAN vs HDBSCAN + <i>Bayesian Optimization</i> ...	82
4.7.	Analisis Hasil Klasterisasi	84
4.8.	Pembuatan GUI.....	89
4.8.1.	Halaman Beranda	89
4.8.2.	Halaman Metode	90
4.8.3.	Halaman Data	90
4.8.4.	Halaman Klasterisasi.....	91
4.8.5.	Halaman Peta Klasterisasi.....	97
BAB V	PENUTUP	99
5.1.	Kesimpulan	99
5.2.	Saran Pengembangan	100
DAFTAR PUSTAKA		103
LAMPIRAN.....		109

DAFTAR GAMBAR

Gambar 2.1. Proses Kerja Metode HDBSCAN.....	23
Gambar 2.2. Alur Kerja <i>Bayesian Optimization</i>	26
Gambar 3.1. Diagram Alir Penelitian	37
Gambar 3.2. <i>Wireframe</i> Aplikasi.....	41
Gambar 3.3. <i>Workflow</i> Sistem Klasterisasi	42
Gambar 4.1. Visualisasi <i>Outlier</i> dengan LOF	54
Gambar 4.2. Visualisasi <i>Scatter Plot</i> HDBSCAN.....	60
Gambar 4.3. Hasil Iterasi <i>Bayesian Optimization</i>	68
Gambar 4.4. Grafik <i>Best DBCV Over Time</i>	70
Gambar 4.5. Grafik Konvergensi <i>Bayesian Optimization</i>	71
Gambar 4.6. Visualisasi <i>Scatter Plot</i> HDBSCAN + <i>Bayesian Optimization</i>	79
Gambar 4.7. Peta Klasterisasi HDBSCAN + <i>Bayesian Optimization</i>	87
Gambar 4.8. Tampilan Halaman Beranda	89
Gambar 4.9. Tampilan Halaman Metode	90
Gambar 4.10. Tampilan Halaman Data	91
Gambar 4.11. Tampilan Hasil Standarisasi Data.....	92
Gambar 4.12. Tampilan Hasil Uji Statistik	92
Gambar 4.13. Tampilan Hasil Deteksi <i>Outlier</i>	93
Gambar 4.14. Tampilan Hasil Reduksi Data	94
Gambar 4.15. Tampilan Pengaturan Rentang Parameter HDBSCAN	94
Gambar 4.16. Tampilan Hasil Pemodelan Klasterisasi	95
Gambar 4.17. Tampilan Tabel Hasil Klasterisasi.....	95
Gambar 4.18. Tampilan Tabel Karakteristik Setiap Klaster	96
Gambar 4.19. Tampilan Interpretasi Hasil Klaster	97
Gambar 4.20. Tampilan Peta Klasterisasi	97
Gambar 4.21. Tampilan Fitur Detail Provinsi dan Unduh Hasil Klaster	98

DAFTAR TABEL

Tabel 2.1. Penelitian Terdahulu.....	7
Tabel 2.2. Kriteria Kelayakan KMO	16
Tabel 3.1. Deskripsi Variabel Penelitian	35
Tabel 3.2. Struktur Dataset.....	36
Tabel 4.1. Dataset Penelitian	45
Tabel 4.2. Statistika Deskriptif Variabel Penelitian	46
Tabel 4.3. Hasil Pemeriksaan <i>Missing Values</i>	48
Tabel 4.4. Hasil Standarisasi Data.....	49
Tabel 4.5. Hasil Perhitungan VIF.....	52
Tabel 4.6. Hasil Perhitungan <i>Eigenvalue</i> dan Proporsi Total Varians	56
Tabel 4.7. Hasil Reduksi PCA.....	57
Tabel 4.8. Nilai <i>Loading Factor</i> Variabel terhadap Komponen Utama.....	58
Tabel 4.9. Contoh Data Implementasi <i>Bayesian Optimization</i>	61
Tabel 4.10. Hasil Perhitungan <i>Expected Improvement</i>	65
Tabel 4.11. 5 Data Awal Hasil Reduksi PCA	72
Tabel 4.12. Hasil Perhitungan Jarak <i>Euclidean</i> 5 Titik Awal	73
Tabel 4.13. Hasil Perhitungan <i>Core Distance</i> 5 Titik Awal	74
Tabel 4.14. Hasil <i>Mutual Reachability Distance</i> 5 Titik Awal	75
Tabel 4.15. Hasil Urutan <i>Edge</i> dan Status Pembentukan Siklus.....	76
Tabel 4.16. Hasil Pembentukan <i>Minimum Spanning Tree</i>	77
Tabel 4.17. Hasil Klaster Perhitungan Manual 5 Titik Awal	78
Tabel 4.18. Hasil Klasterisasi HDBSCAN + <i>Bayesian Optimization</i>	80
Tabel 4.19. Hasil Evaluasi DBSCAN vs HDBSCAN + <i>Bayesian Optimization</i> .	81
Tabel 4.20. Perbandingan HDBSCAN vs HDBSCAN + <i>Bayesian Optimization</i>	82
Tabel 4.21. Rata-Rata Variabel Setiap Klaster.....	84
Tabel 4.22. Persentase Kontribusi Variabel Setiap Klaster.....	84

DAFTAR LAMPIRAN

Lampiran 1. Dataset Penelitian	109
Lampiran 2. Kode Script Penelitian	110
Lampiran 3. Kode Script GUI	111
Lampiran 4. LoA Publikasi	112

DAFTAR NOTASI

n	:	Jumlah observasi
p	:	Jumlah variabel
μ	:	Rata-rata
σ	:	Standar deviasi
r_{ij}^2	:	Kuadrat koefisien korelasi antara variabel i dan j
a_{ij}^2	:	Kuadrat koefisien korelasi parsial antara variabel i dan j
χ^2	:	Nilai statistik uji <i>chi-square</i>
$\ R\ $	:	Determinan matriks korelasi
VIF_i	:	Nilai <i>Variance Inflation Factor</i> variabel ke- i
R_i^2	:	Koefisien determinasi variabel ke- i terhadap variabel independen lainnya
$cov(X)$	:	Matriks kovarians
$N_k(a)$	:	Jumlah tetangga titik a
$lrd_k(a)$	:	Nilai <i>local reachability density</i> titik a
$reach_dist_k(a, b)$	:	Nilai <i>reachability distance</i> dari titik a ke titik b
$LOF_k(a)$	:	Nilai <i>Local Outlier Factor</i> titik p
ε	:	Radius <i>epsilon</i>
$d(a, b)$	:	Jarak <i>Euclidean</i> titik a dan b
$core_k(a)$	:	Jarak inti (<i>core distance</i>) titik a ke tetangga ke- k terdekat
$d_{mreach-k}(a, b)$	:	Jarak keterjangkauan bersama (<i>Mutual Reachability Distance</i>) antara titik a dan b
$\mu_n(x)$	:	Nilai <i>mean</i> prediktif <i>Gaussian Process</i> pada titik x
$\sigma^2(x)$	:	Nilai <i>variance</i> prediktif <i>Gaussian Process</i> pada titik x
$EI(x)$	:	Nilai <i>Expected Improvement</i> pada titik x