

MODIFIKASI JARAK DAFI-GOWER UNTUK DATA CAMPURAN YANG MENGANDUNG *OUTLIER* PADA PENGGEROMBOLAN KECAMATAN DI JAWA BARAT

DELIA FITRI AUDINA



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



@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 “Modifikasi Jarak DAFI-Gower untuk Data Campuran yang Mengandung *Outlier* pada Penggerombolan Kecamatan 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.

Dalam penyusunan karya ini, penulis menggunakan bantuan kecerdasan buatan berupa ChatGPT dan Claude dalam pencarian referensi ilmiah, komplikasi dan perbaikan kode analisis, proses brainstorming, serta penyuntingan dan perbaikan redaksi dalam penulisan skripsi. Setelah menggunakan alat/layanan tersebut, penulis meninjau dan menyunting konten sesuai kebutuhan serta bertanggungjawab penuh atas isi karya tugas akhir ini.

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

Bogor, Juli 2026

Delia Fitri Audina
G1401221013



ABSTRAK

DELIA FITRI AUDINA. Modifikasi Jarak DAFI-Gower untuk Data Campuran yang Mengandung *Outlier* pada Penggerombolan Kecamatan di Jawa Barat. Dibimbing oleh ANWAR FITRIANTO dan BUDI SUSETYO.

Analisis gerombol pada data campuran memerlukan ukuran jarak yang mampu mengakomodasi peubah numerik dan kategorik secara simultan. Keberadaan *outlier* menyebabkan pemilihan ukuran penyebaran pada peubah numerik menjadi penting karena dapat memengaruhi kualitas perhitungan jarak dan hasil penggerombolan. Penelitian ini bertujuan memodifikasi ukuran jarak DAFI-Gower menggunakan *Standard Deviation* (SD) dan *Median Absolute Deviation* (MAD) sebagai ukuran penyebaran pada peubah numerik, membandingkan kualitas hasil penggerombolan menggunakan metode *K-Prototypes*, *Partitioning Around Medoids* (PAM) berbasis jarak Gower, PAM berbasis DAFI-Gower, serta PAM berbasis modifikasi DAFI-Gower, dan mengidentifikasi karakteristik gerombol berdasarkan metode terbaik. Penelitian ini menggunakan data sekunder yang terdiri atas 627 kecamatan di Provinsi Jawa Barat dengan 8 peubah numerik dan 4 peubah kategorik. Kualitas penggerombolan dievaluasi menggunakan koefisien *Silhouette*, *Davies-Bouldin Index* (DBI), dan *Dunn Index*. Hasil penelitian menunjukkan bahwa PAM berbasis modifikasi DAFI-Gower dengan MAD menghasilkan kualitas penggerombolan terbaik dengan nilai koefisien *Silhouette* sebesar 0,520 dan DBI sebesar 1,747. Metode tersebut menghasilkan dua gerombol kecamatan dengan karakteristik pembangunan yang berbeda. Temuan penelitian menunjukkan bahwa penggunaan MAD sebagai ukuran penyebaran pada DAFI-Gower mampu meningkatkan kualitas penggerombolan data campuran yang mengandung *outlier* sehingga dapat menjadi alternatif modifikasi ukuran jarak yang lebih *robust* untuk analisis data campuran.

Kata kunci: DAFI-Gower, data campuran, *median absolute deviation*, *outlier*, PAM.



ABSTRACT

DELIA FITRI AUDINA. Modification of the DAFI-Gower Distance for Mixed Data Containing Outliers in the Clustering of Subdistricts in West Java. Supervised by ANWAR FITRIANTO and BUDI SUSETYO.

Cluster analysis of mixed data requires a distance measure that can simultaneously accommodate numerical and categorical variables. The presence of outliers makes the selection of an appropriate dispersion measure for numerical variables important because it may affect distance calculation and clustering performance. This study aimed to modify the DAFI-Gower distance by using Standard Deviation (SD) and Median Absolute Deviation (MAD) as dispersion measures, compare the clustering performance of K-Prototypes, Partitioning Around Medoids (PAM) based on the Gower distance, DAFI-Gower distance, and the modified DAFI-Gower distance, and identify cluster characteristics using the best-performing method. Secondary data from 627 subdistricts in West Java Province with eight numerical and four categorical variables were analyzed. Clustering performance was evaluated using the Silhouette Coefficient, Davies–Bouldin Index (DBI), and Dunn Index. The results showed that PAM based on the modified DAFI-Gower distance using MAD achieved the best clustering performance, with a Silhouette Coefficient of 0.520 and a DBI of 1.747. The method produced two clusters of subdistricts with distinct development characteristics. The findings indicate that using MAD as the dispersion measure in the DAFI-Gower distance improves clustering performance for mixed data containing outliers, providing a more robust alternative for mixed data clustering.

Keywords: DAFI-Gower, median absolute deviation, mixed data, outlier, PAM.



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 2026¹
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.

MODIFIKASI JARAK DAFI-GOWER UNTUK DATA CAMPURAN YANG MENGANDUNG *OUTLIER* PADA PENGGEROMBOLAN KECAMATAN DI JAWA BARAT

DELIA FITRI AUDINA

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
2026**



@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. I Made Sumertajaya, M.Si.



@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.

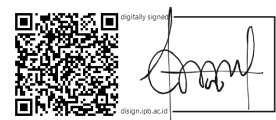


Judul Skripsi : Modifikasi Jarak DAFI-Gower untuk Data Campuran yang
Mengandung *Outlier* pada Penggerombolan Kecamatan di Jawa
Barat

Nama : Delia Fitri Audina
NIM : G1401221013

Disetujui oleh

Pembimbing 1:
Dr. Anwar Fitrianto, S.Si., M.Sc.



Pembimbing 2:
Dr. Ir. Budi Susetyo, M.S.



Diketahui oleh

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

Tanggal Ujian:
17 Juli 2026

Tanggal Lulus:

PRAKATA

Puji dan syukur penulis panjatkan kepada Allah SWT atas segala rahmat, karunia, serta pertolongan-Nya sehingga karya ilmiah yang berjudul “Modifikasi Jarak DAFI Gower untuk Data Campuran yang Mengandung *Outlier* pada Penggerombolan Kecamatan di Jawa Barat” dapat diselesaikan dengan baik. Penyusunan karya ilmiah ini tidak terlepas dari bantuan, doa, bimbingan, dukungan, serta motivasi dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis menyampaikan terima kasih kepada:

1. Bapak Dr. Anwar Fitrianto, S.Si., M.Sc. dan Bapak Dr. Ir. Budi Susetyo, M.S. selaku dosen pembimbing yang telah meluangkan waktu, tenaga, dan pikiran untuk memberikan arahan, bimbingan, masukan, serta motivasi kepada penulis dengan penuh kesabaran selama proses penyusunan karya ilmiah ini.
2. Ibu Laily Nissa Atul Mualifah, S.Si, M.Si., Bapak Dr. Farit Mochamad Afendi, S.Si., M.Si., serta Bapak Prof. Dr. Ir. I Made Sumertajaya, M.Si selaku moderator kolokium, seminar hasil, dan penguji luar yang juga telah memberikan kritik, saran, serta masukan yang membangun sehingga karya ilmiah ini dapat diselesaikan dengan lebih baik.
3. Kedua orang tua tercinta, khususnya Ibu Lilis Saodah yang senantiasa memberikan doa, kasih sayang, dukungan, perhatian, serta semangat kepada penulis dalam setiap langkah kehidupan. Penulis juga mengenang Bapak A. Suhara (Alm.) yang selalu menjadi motivasi, inspirasi, dan sumber kekuatan bagi penulis untuk terus berjuang menyelesaikan pendidikan ini. Ucapan terima kasih juga penulis sampaikan kepada seluruh keluarga besar atas doa, perhatian, dan dukungan yang tidak pernah putus.
4. Salsabila, Raihana, dan Salsa yang telah menjadi sahabat terdekat selama masa perkuliahan, selalu memberikan semangat, motivasi, dukungan, serta menemani penulis dalam menghadapi berbagai tantangan selama menjalani perkuliahan hingga penyusunan karya ilmiah ini.
5. Cindy, Bulan, Vito, Jidan yang telah menjadi sahabat sejak masa pengenalan jurusan hingga saat ini, serta selalu berbagi kebersamaan, pengalaman, semangat, dan dukungan sehingga masa perkuliahan menjadi lebih bermakna dan penuh kenangan.
6. Seluruh keluarga besar Marinestic Statistika 59 atas kebersamaan, kerja sama, pengalaman, serta berbagai kenangan yang telah diberikan selama menjalani perkuliahan.
7. Semua pihak yang tidak dapat disebutkan satu per satu yang telah memberikan bantuan, doa, dukungan, maupun kontribusi, baik secara langsung maupun tidak langsung, dalam penyelesaian karya ilmiah ini.

Semoga karya ilmiah ini dapat memberikan manfaat bagi pembaca dan pengembangan ilmu pengetahuan, khususnya di bidang statistika.

Bogor, Juli 2026

Delia Fitri Audina



DAFTAR ISI

DAFTAR TABEL	xi
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xii
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan	3
TINJAUAN PUSTAKA	4
2.1 Analisis Gerombol (<i>Cluster Analysis</i>)	4
2.2 <i>Outlier</i>	4
2.3 <i>Adjusted Boxplot</i>	4
2.4 Jarak Gower	5
2.5 DAFI-Gower	6
2.6 Ukuran Penyebaran Data	8
2.6.1 <i>Interquartile Range (IQR)</i>	8
2.6.2 <i>Standard Deviation (SD)</i>	8
2.6.3 <i>Median Absolute Deviation (MAD)</i>	9
2.7 <i>Partitioning Around Medoids (PAM)</i>	9
2.8 <i>K-Prototypes</i>	9
2.9 Koefisien <i>Silhouette</i>	10
2.10 <i>Davies-Bouldin Index (DBI)</i>	11
2.11 <i>Dunn Index</i>	11
III METODE	13
3.1 Data	13
3.2 Prosedur Analisis Data	13
IV HASIL DAN PEMBAHASAN	16
4.1 Praproses Data	16
4.2 Eksplorasi Data	16
4.3 Hasil Penggerombolan Metode <i>K-Prototypes</i>	19
4.4 Hasil Penggerombolan Metode PAM Berbasis Jarak Gower	19
4.5 Hasil Penggerombolan Metode PAM Berbasis DAFI-Gower	21
4.5.1 Pembobotan <i>Feature Importance</i> dan <i>Distance Adjustment</i>	21
4.5.2 Hasil Penggerombolan	23
4.6 Perbandingan Hasil Penggerombolan	25
4.7 Analisis Keanggotaan Gerombol	26
4.8 Karakteristik Gerombol Hasil Penggerombolan Terbaik	27
V SIMPULAN DAN SARAN	30
5.1 Simpulan	30
5.2 Saran	30
DAFTAR PUSTAKA	31
LAMPIRAN	33
RIWAYAT HIDUP	39



DAFTAR TABEL

1	Interpretasi nilai koefisien <i>Silhouette</i>	11
2	Peubah yang digunakan dalam penelitian	13
3	Distribusi anggota gerombol metode <i>K-Prototypes</i>	19
4	Distribusi anggota gerombol metode PAM berbasis jarak Gower	20
5	Nilai <i>distance adjustment</i> setiap peubah	22
6	Distribusi anggota gerombol PAM berbasis DAFI-Gower (IQR)	23
7	Distribusi anggota gerombol PAM berbasis DAFI-Gower (SD)	25
8	Distribusi anggota gerombol PAM berbasis DAFI-Gower (MAD)	25
9	Hasil evaluasi metode penggerombolan	26
10	Kesesuaian keanggotaan gerombol antarmetode	27
11	Karakteristik setiap gerombol hasil penggerombolan	28

DAFTAR GAMBAR

1	Prosedur analisis data	15
2	<i>Adjusted boxplot</i> peubah numerik: (a) kepadatan penduduk, (b) jarak ke ibukota, (c) rasio sarana pendidikan per 1000, (d) rasio sarana kesehatan per 1000, (e) rasio posyandu aktif per 1000, (f) rasio sarana perdagangan per 1000, (g) rasio koperasi per 1000, dan (h) rasio BTS per 1000	16
3	<i>Bar chart</i> peubah kategorik: (a) akses angkutan umum, (b) sumber penghasilan utama, (c) kekuatan sinyal, dan (d) penerangan jalan	18
4	Penentuan jumlah gerombol optimal metode <i>K-Prototypes</i> berdasarkan koefisien <i>Silhouette</i>	19
5	Penentuan jumlah gerombol optimal metode PAM berbasis jarak Gower berdasarkan koefisien <i>Silhouette</i>	20
6	Bobot <i>feature importance</i> untuk setiap peubah	21
7	Penentuan jumlah gerombol optimal metode PAM berbasis DAFI-Gower (IQR) berdasarkan koefisien <i>Silhouette</i>	23
8	Penentuan jumlah gerombol optimal metode PAM berbasis DAFI-Gower (SD) berdasarkan koefisien <i>Silhouette</i>	24
9	Penentuan jumlah gerombol optimal metode PAM berbasis DAFI-Gower (MAD) berdasarkan koefisien <i>Silhouette</i>	25
10	Peta hasil penggerombolan menggunakan metode PAM berbasis DAFI-Gower (MAD)	29



DAFTAR LAMPIRAN

1	Definisi operasional peubah numerik	34
2	Komponen penyusun beberapa peubah numerik	34
3	Definisi operasional dan kategori peubah kategorik	35
4	Jumlah <i>outlier</i> pada peubah numerik	35
5	Matriks <i>Normalized Mutual Information</i> (NMI) antarpeubah	36
6	Bobot <i>feature importance</i> setiap peubah	37
7	Nilai DXj peubah kategorik	37
8	Nilai Qj untuk IQR, SD, dan MAD	37
9	Visualisasi karakteristik gerombol berdasarkan metode penggerombolan terbaik	38

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.