

**PENGELOMPOKAN PROVINSI DI INDONESIA
BERDASARKAN JUMLAH KASUS KRIMINALITAS DENGAN
MENGUNAKAN METODE *CLUSTERING K-MEDOIDS***

SKRIPSI



Oleh :
ZULKHAIRI FAHMI.Z
2117010032

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UIN IMAM BONJOL PADANG
2025**

**PENGELOMPOKAN PROVINSI DI INDONESIA
BERDASARKAN JUMLAH KASUS KRIMINALITAS DENGAN
MENGUNAKAN METODE *CLUSTERING K-MEDOIDS***

SKRIPSI

Diajukan Sebagai Salah satu Syarat untuk Meraih Gelar Sarjana (S1)



**UIN IMAM BONJOL
PADANG**

**Oleh :
ZULKHAIRI FAHMI.Z
2117010032**

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UIN IMAM BONJOL PADANG
2025**

HALAMAN PERNYATAAN ORISINALITAS

**Skripsi ini adalah hasil karya saya sendiri,
dan semua sumber baik yang dikutip maupun dirujuk
telah saya nyatakan dengan benar.**

Padang, 02 Juli 2025

Yang Membuat Pernyataan,



ZULKHAIRI FAHMI.Z

HALAMAN PERSETUJUAN

Skripsi dengan judul **Pengelompokan Provinsi di Indonesia Berdasarkan Jumlah Kasus Kriminalitas dengan Menggunakan Metode *Clustering K-Medoids*** yang disusun oleh **Zulkhairi Fahmi.Z** NIM 2117010032, telah diperiksa dan disetujui untuk diajukan dalam sidang skripsi pada Program Studi Matematika Fakultas Sains dan Teknologi UIN Imam Bonjol Padang.

Padang, 23 April 2025

Menyetujui:

Pembimbing I



Syarto Musthofa, M.Sc.
NIP 199309152020121012

Pembimbing II

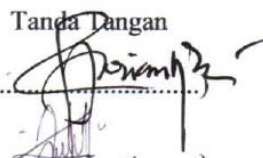
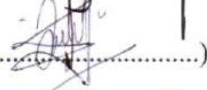
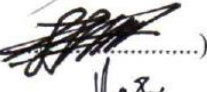


Darvi Mailisa Putri, M.Si.
NIP 199006162019032021

HALAMAN PENGUJIAN

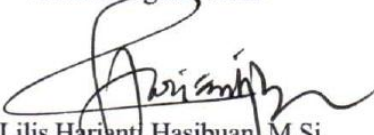
Skripsi dengan judul **Pengelompokan Provinsi di Indonesia Berdasarkan Jumlah Kasus Kriminalitas dengan Menggunakan Metode *Clustering K-Medoids*** yang disusun oleh **Zulkhairi Fahmi.Z** NIM 2117010032, telah diuji di depan Tim Penguji pada hari Rabu tanggal 14 Mei 2025.

Tim Penguji:

	Nama	Jabatan	Tanda Tangan
1	Lilis Harianti Hasibuan, M.Si. NIP 198804142019032008	(Ketua)	(..... )
2	Ilham Dangu Rianjaya, M.Si. NIP 199207022020121009	(Sekretaris)	(..... )
3	Syarto Musthofa, M.Sc. NIP 199309152020121012	(Anggota)	(..... )
4	Darvi Mailisa Putri, M.Si. NIP 199006162019032021	(Anggota)	(..... )

Mengetahui:

Ketua Program Studi


Lilis Harianti Hasibuan, M.Si.
NIP 198804142019032008

HALAMAN PENGESAHAN

Skripsi dengan judul **Pengelompokan Provinsi di Indonesia Berdasarkan Jumlah Kasus Kriminalitas dengan Menggunakan Metode *Clustering K-Medoids*** yang disusun oleh **Zulkhairi Fahmi.Z** NIM 2117010032 Program Studi Matematika Fakultas Sains dan Teknologi UIN Imam Bonjol Padang telah dipertahankan di depan Tim Penguji Sidang Skripsi pada tanggal 14 Mei 2025 dan dinyatakan memenuhi syarat.

Disahkan di : Padang

Tanggal : 02 Juli 2025

Menyetujui:

Pembimbing I



Syarto Musthofa, M.Sc.
NIP 199309152020121012

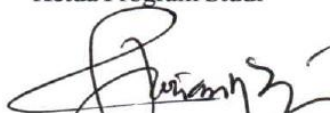
Pembimbing II



Darvi Mailisa Putri, M.Si.
NIP 199006162019032021

Mengetahui,

Ketua Program Studi



Lilis Harianti Hasibuan, M.Si.
NIP 198804142019032008

Mengesahkan,

Dekan Fakultas Sains dan Teknologi




Prof. Nurus Shalihin, M.Si., Ph.D.
NIP 196911192003121001

KATA PENGANTAR

Puji syukur kepada Allah SWT karena berkat rahmat dan karunia-Nya sehingga penyusunan skripsi dengan judul “Pengelompokan Provinsi di Indonesia Berdasarkan Jumlah Kasus Kriminalitas dengan Menggunakan Metode *Clustering K-Medoids*” dapat diselesaikan dengan baik. Penulis menyadari bahwa selama penyusunan skripsi, penulis dihadapkan dengan berbagai hambatan dan kendala, namun berkat Allah SWT, bantuan berbagai pihak dan kerja keras akhirnya penyusunan skripsi ini dapat diselesaikan. Skripsi ini disusun sebagai salah satu syarat untuk menyelesaikan pendidikan di Program Studi Matematika Fakultas Sains dan Teknologi UIN Imam Bonjol Padang.

Dalam penulisan skripsi ini tentunya masih jauh dari kata sempurna, hal ini dikarenakan keterbatasan pengetahuan yang penulis miliki. Oleh karena itu, dalam melengkapi kesempurnaan untuk penulisan skripsi ini diharapkan adanya saran dan kritik yang diberikan bersifat membangun. Pada kesempatan yang baik ini tak luput pula penulis menghaturkan banyak terima kasih kepada semua pihak yang telah memberikan bimbingan, pengarahan, nasehat, dan pemikiran dalam penulisan skripsi ini, terutama kepada :

1. Dekan dan Civitas Akademia Fakultas Sains dan Teknologi.
2. Ketua Program Studi Matematika dan Sekretaris Program Studi Matematika Fakultas Sains dan Teknologi.
3. Syarto Musthofa, M.Sc. selaku Pembimbing I yang telah memberikan arahan dalam penulisan skripsi ini.
4. Darvi Mailisa Putri, M.Si. selaku dosen Pembimbing II yang telah memberikan arahan dalam penulisan skripsi ini.
5. Syarto Musthofa, M.Sc. selaku Pembimbing Akademik yang telah memberikan arahan selama masa perkuliahan.
6. Lilis Harianti Hasibuan, M.Si. selaku dosen penguji dalam seminar proposal yang telah memberikan masukan dalam penulisan proposal skripsi.

7. Seluruh dosen Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Imam Bonjol Padang yang telah memberikan dukungan, semangat dan arahan dalam penyelesaian penulisan skripsi.
8. Kedua orang tua penulis, yang dengan segenap cinta, kesabaran, dan pengorbanan telah membesarkan, membimbing, serta mendampingi penulis dalam setiap tahap kehidupan. Dalam setiap langkah yang penulis tempuh, terdapat doa tulus yang tidak pernah henti dari kedua orang tua. Setiap tetes keringat, kerja keras, dan keikhlasan yang diberikan menjadi pondasi kokoh bagi pencapaian ini. Tanpa kasih sayang yang tak ternilai, serta bimbingan moral dan spiritual yang terus mengiringi, penulis tidak akan mampu sampai pada titik ini. Skripsi ini penulis persembahkan sebagai wujud kecil dari rasa hormat, cinta, dan terima kasih yang mendalam atas segala yang telah diberikan selama ini.
9. Saudara kandung penulis, yaitu Ihsanul Fuadi Z dan Muhammad Irsyad Z yang telah memberikan semangat dan menjadi salah satu pondasi utama keteguhan diri penulis dalam proses penyelesaian skripsi ini.
10. Seluruh mahasiswa matematika seperjuangan yang saling menyemangati dalam menyelesaikan studi di kampus ini.
11. Kepada semua pihak yang telah meluangkan waktunya untuk membantu dalam penyelesaian skripsi ini yang tidak bisa disebutkan satu persatu.

Penulis menyadari skripsi ini tidak luput dari berbagai kekurangan. Penulis mengharapkan saran dan kritik yang membangun demi kesempurnaan dan perbaikannya sehingga laporan skripsi ini dapat memberikan manfaat bagi bidang pendidikan dan bisa dikembangkan lebih lanjut lagi. Aamiin.

Padang, 02 Juli 2025



Penulis

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI KARYA
ILMIAH UNTUK KEPENTINGAN AKADEMIS**

Saya yang bertanda tangan di bawah ini:

Nama : Zulkhairi Fahmi.Z
NIM : 2117010032
Program Studi : Matematika
Judul Skripsi : Pengelompokan Provinsi di Indonesia Berdasarkan
Jumlah Kasus Kriminalitas dengan Menggunakan
Metode *Clustering K-Medoids*

Dengan ini menyatakan persetujuan untuk dipublikasikannya karya ilmiah (skripsi)
saya oleh pihak UIN Imam Bonjol Padang untuk kepentingan akademis.

Padang, 02 Juli 2025

Yang Membuat Pernyataan,

A 10,000 Rupiah Indonesian postage stamp is placed over the signature. The stamp features the Garuda Pancasila emblem, the text 'SEPULUH RIBU RUPIAH', '10000', and 'METERAI TEMPEL'. The serial number '4068AMX403715694' is visible at the bottom of the stamp.

Zulkhairi Fahmi.Z

ABSTRAK

Kriminalitas merupakan salah satu masalah sosial yang paling mempengaruhi keamanan dan kesejahteraan masyarakat di suatu daerah. Perbedaan jumlah kasus kriminalitas yang relatif fluktuatif antar wilayah di Indonesia menjadi perhatian yang perlu diidentifikasi. Penelitian ini bertujuan untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan jumlah kasus kriminalitas dengan menggunakan metode pengelompokan *K-Medoids*. Data yang digunakan pada penelitian ini adalah jumlah kasus-kasus kriminalitas pada setiap provinsi di Indonesia. Terdapat tiga pengukuran jarak yang digunakan, yaitu *Euclidean Distance*, *Manhattan Distance*, dan *Minkowski Distance*. Kualitas pengelompokan akan dievaluasi dengan menggunakan *Davies-Bouldin Index* dan *Silhouette Score*. Hasil penelitian menunjukkan bahwa pengelompokan menggunakan metode *K-Medoids* dengan pengukuran jarak *Euclidean Distance* dan *Minkowski Distance* memberikan kualitas pengelompokan yang lebih baik dibandingkan dengan pengelompokan yang menggunakan pengukuran jarak *Manhattan Distance* berdasarkan *Silhouette Score* dan *Davies-Bouldin Index* yaitu dengan nilai 0.746 dan 0.42 secara berturut-turut, sehingga pengelompokan tersebut menjadi solusi pada penelitian.

Kata kunci: Kriminalitas, Metode *K-Medoids*, Pengelompokan

ABSTRACT

Crime is one of the social problems that most affects the security and welfare of the community in a region. The difference in the number of crime cases that are relatively fluctuating between regions in Indonesia is a concern that needs to be identified. This research aims to group provinces in Indonesia based on the number of crime cases using the K-Medoids clustering method. The data used in this research is the number of crime cases in each province in Indonesia. There are three distance measurements used, namely Euclidean Distance, Manhattan Distance, and Minkowski Distance. The quality of clustering will be evaluated using Davies-Bouldin Index and Silhouette Score. The results show that clustering using the K-Medoids method with Euclidean Distance and Minkowski Distance distance measurements provides better clustering quality compared to clustering using Manhattan Distance distance measurements based on Silhouette Score and Davies-Bouldin Index with values of 0.746 and 0.42 respectively, so that the clustering becomes the solution in the study.

Keywords: *Crime, K-Medoids Method, Clustering*

DAFTAR ISI

HALAMAN PERNYATAAN ORISINALITAS	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGUJIAN	iv
HALAMAN PENGESAHAN.....	v
KATA PENGANTAR.....	vi
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS	viii
ABSTRAK	ix
DAFTAR ISI.....	xi
DAFTAR TABEL.....	xiii
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	5
BAB II TINJAUAN PUSTAKA.....	6
2.1 Kerangka Teori.....	6
2.1.1 Kriminalitas.....	6
2.1.2 Metode Pengukuran Jarak	8
2.1.3 <i>Dunn Index</i> Menentukan <i>Cluster</i> Optimum.....	10
2.1.4 <i>Clustering</i>	10
2.1.5 <i>Outlier</i>	13
2.1.6 Evaluasi Kualitas <i>Clustering</i>	15
2.2 Literatur Review	17
BAB III METODOLOGI PENELITIAN	19
3.1 Jenis Penelitian.....	19
3.2 Bahan Penelitian.....	19

3.3 Variabel Penelitian	19
3.4 Teknik Pengumpulan Data	20
3.5 Tahapan Penelitian	21
BAB IV HASIL DAN PEMBAHASAN.....	22
4.1 Statistik Deskriptif	22
4.2 Deteksi <i>Outlier</i>	23
4.3 Penentuan Jumlah <i>Cluster</i> Optimum dengan <i>Dunn Index</i>	26
4.4 Proses Pengelompokan dengan Metode <i>Clustering K-Medoids</i>	27
4.4.1 Pengelompokan dengan <i>Euclidean Distance</i>	27
4.4.2 Pengelompokan dengan <i>Manhattan Distance</i>	32
4.4.3 Pengelompokan dengan <i>Minkowski Distance</i>	36
4.5 Evaluasi Kualitas Pengelompokan	42
4.5.1 <i>Silhouette Score</i>	42
4.5.2 <i>Davies-Bouldin Index</i>	47
4.6 Interpretasi Hasil Pengelompokan dengan Kualitas Terbaik	50
BAB V PENUTUP	54
5.1 Kesimpulan	54
5.2 Saran	55
DAFTAR PUSTAKA	56
LAMPIRAN	62

DAFTAR TABEL

Tabel 2.1.	Rentang Nilai dan Interpretasi <i>Silhouette Score</i>	15
Tabel 3.1.	Data Yang Digunakan pada Penelitian.....	19
Tabel 4.1.	Statistik Deskriptif Keseluruhan Variabel.....	22
Tabel 4.2.	Matriks Varians Kovarians Data	24
Tabel 4.3.	Nilai <i>Mahalanobis Distance</i>	24
Tabel 4.4.	Nilai <i>Dunn Index</i>	27
Tabel 4.5.	<i>Medoids</i> pada Iterasi ke-1 Pengelompokan <i>Euclidean Distance</i>	28
Tabel 4.6.	<i>Medoids</i> pada Iterasi ke-2 Pengelompokan <i>Euclidean Distance</i>	29
Tabel 4.7.	Solusi <i>Medoids</i> pada Pengelompokan dengan <i>Euclidean Distance</i>	30
Tabel 4.8.	Pengelompokan dengan Pengukuran Jarak <i>Euclidean Distance</i>	31
Tabel 4.9.	<i>Medoids</i> pada Iterasi ke-1 Pengelompokan <i>Manhattan Distance</i>	33
Tabel 4.10.	<i>Medoids</i> pada Iterasi ke-2 Pengelompokan <i>Manhattan Distance</i>	33
Tabel 4.11.	Solusi <i>Medoids</i> dari <i>Software Rstudio</i> Pengelompokan <i>Manhattan Distance</i>	34
Tabel 4.12.	Pengelompokan dengan Pengukuran Jarak <i>Manhattan Distance</i>	35
Tabel 4.13.	Solusi <i>Medoids</i> dengan <i>Minkowski Distance</i> pada $p = 3$	36
Tabel 4.14.	Pengelompokan dengan Pengukuran Jarak <i>Minkowski Distance</i> pada $p = 3$	37
Tabel 4.15.	Solusi <i>Medoids</i> dengan <i>Minkowski Distance</i> pada $p = 4$ dari <i>Software Rstudio</i>	38
Tabel 4.16.	Pengelompokan dengan Pengukuran Jarak <i>Minkowski Distance</i> pada $p = 4$	38

Tabel 4.17. Solusi Medoids dengan Minkowski Distance pada $p = 5$ dari <i>Software Rstudio</i>	40
Tabel 4.18. Pengelompokan dengan Pengukuran Jarak <i>Minkowski</i> <i>Distance</i> pada $p = 5$	40
Tabel 4.19. Hasil Pengelompokan.....	51
Tabel 4.20. Rata-Rata Variabel Setiap <i>Cluster</i>	52
Tabel 4.21. Hasil Akhir Pengelompokan Metode K-Medoids	53

DAFTAR LAMPIRAN

Lampiran 1. Nilai Titik Data Tiap Variabel Dikurangi Vektor Rata-Ratanya.....	62
Lampiran 2. Matriks Varians Kovarians <i>Mahalanobis Distance</i>	64
Lampiran 3. Jarak Titik Data ke Medoids Iterasi ke-1 Pengelompokan dengan <i>Euclidean Distance</i>	65
Lampiran 4. Nilai Cost Tiap Data dan Total Cost Iterasi ke-1 Pengelompokan dengan <i>Euclidean Distance</i>	66
Lampiran 5. Jarak Titik Data ke <i>Medoids</i> , Nilai <i>Cost</i> Tiap Data dan Nilai Total <i>Cost</i> pada Iterasi ke-2 Pengelompokan <i>Euclidean Distance</i>	67
Lampiran 6. Iterasi dari <i>Medoids</i> yang Menjadi Solusi Pengelompokan <i>K-Medoids</i> dengan <i>Euclidean Distance</i>	68
Lampiran 7. Jarak Titik Data ke <i>Medoids</i> Iterasi ke-1 Pengelompokan dengan <i>Manhattan Distance</i>	69
Lampiran 8. Nilai <i>Cost</i> Tiap Data dan Total Cost Iterasi ke-1 Pengelompokan dengan <i>Manhattan Distance</i>	70
Lampiran 9. Jarak Titik Data ke <i>Medoids</i> , Nilai <i>Cost</i> Tiap Data dan Nilai Total <i>Cost</i> pada Iterasi ke-2 Pengelompokan <i>Manhattan Distance</i>	71
Lampiran 10. Iterasi dari <i>Medoids</i> yang Menjadi Solusi Pengelompokan <i>K-Medoids</i> dengan <i>Manhattan Distance</i>	72
Lampiran 11. Nilai a_i <i>Silhouette Score</i> Keseluruhan Titik Data Pengelompokan dengan <i>Euclidean Distance</i>	73
Lampiran 12. Jarak Rata-Rata antara Keseluruhan Titik Data antara Pengelompokan <i>Euclidean Distance</i> dengan Titik Data Lain pada <i>Cluster 1</i> dan <i>Cluster 2</i>	74
Lampiran 13. Nilai b_i <i>Silhouette Score</i> Keseluruhan Titik Data Pengelompokan dengan <i>Euclidean Distance</i>	75

Lampiran 14. Nilai <i>Silhouette Score</i> Keseluruhan Titik Data Pengelompokan dengan <i>Euclidean Distance</i>	76
Lampiran 15. Nilai a_i dan b_i <i>Silhouette Score</i> Keseluruhan Titik Data Pengelompokan dengan <i>Manhattan Distance</i>	77
Lampiran 16. Nilai <i>Silhouette Score</i> Keseluruhan Titik Data Pengelompokan dengan <i>Manhattan Distance</i>	78
Lampiran 17. Nilai a_i , b_i , dan <i>Silhouette Score</i> Keseluruhan Titik Data Pengelompokan dengan <i>Minkowski Distance</i>	79
Lampiran 18. Nilai Jarak dari Tiap-Tiap Pasang <i>Centroid</i> (b_i, j) <i>Davies-Bouldin Index</i> Pengelompokan <i>Euclidean Distance</i>	80
Lampiran 19. Nilai Jarak Rata-Rata Semua Titik Data pada Suatu Cluster dengan <i>Centroid Cluster</i> untuk Setiap <i>Cluster</i> ke- i (avg_i) Pengelompokan <i>Euclidean Distance</i>	81
Lampiran 20. Nilai Similitude antara Dua <i>Cluster</i> yang Berdekatan ($R_{i,j}$) Pengelompokan dengan <i>Euclidean Distance</i>	82
Lampiran 21. Nilai <i>Davies-Bouldin Index</i> untuk Setiap <i>Cluster</i> (DB_i) Pengelompokan dengan <i>Euclidean Distance</i>	83
Lampiran 22. Nilai $d_{i,j}$, avg_i , $R_{i,j}$, dan DB_i <i>Davies-Bouldin Index</i> Pengelompokan dengan <i>Manhattan Distance</i>	84
Lampiran 23. Nilai $d_{i,j}$, avg_i , $R_{i,j}$, dan DB_i <i>Davies-Bouldin Index</i> Pengelompokan dengan <i>Minkowski Distance</i>	85
Lampiran 24. <i>Syntax Software Rstudio</i> untuk <i>Clustering K-Medoids</i>	86
Lampiran 25. <i>Syntax Software Rstudio</i> untuk Menghitung Nilai Jarak <i>Mahalanobis Distance</i>	87
Lampiran 26. <i>Syntax Software Rstudio</i> untuk Menghitung Nilai <i>Silhouette Score</i>	88