

KLASTERISASI DAERAH KRIMINALITAS DI INDONESIA DENGAN METODE K-MEANS CLUSTERING

T.Sofia Chairani^{1*}, Hijka Listia², Sabina Wardaniah³, Siti Wulandari⁴, Putri Tasya Agustina Tampubolon⁵, Arnita Piliang⁶

^{1,2,3,4,5,6}Universitas Negeri Medan

Email: ^{1*}tsofiachairani@gmail.com, ²hsbtia62@mhs.unimed.ac.id, ³sabinawardaniah@mhs.unimed.ac.id, ⁴sitiwulandari271023@gmail.com, ⁵ciyatampu@gmail.com, ⁶arnita@unimed.ac.id

(Naskah masuk: 3 September 2024, diterima untuk diterbitkan: 30 September 2024)

Abstrak

Kriminalitas di Indonesia menjadi masalah signifikan yang memengaruhi berbagai aspek seperti keamanan, sosial, dan ekonomi. Namun, upaya mitigasi sering terkendala oleh kurangnya informasi yang terstruktur tentang daerah rawan kriminalitas. Penelitian ini bertujuan untuk mengatasi permasalahan tersebut dengan mengelompokkan provinsi di Indonesia berdasarkan pola kriminalitas menggunakan algoritma K-Means Clustering. Data statistik kriminal tahun 2014-2023 dari Badan Pusat Statistik (BPS) dianalisis dengan menentukan jumlah cluster optimal yang menghasilkan lima cluster dengan evaluasi menggunakan Python dan library Scikit-learn. Hasil penelitian menunjukkan Silhouette Score sebesar 0.593, yang mencerminkan pembentukan cluster cukup baik. Klasterisasi ini memberikan panduan berbasis data bagi pemerintah dalam menyusun kebijakan keamanan yang lebih terarah untuk mengurangi tingkat kriminalitas di Indonesia.

Kata kunci: kriminalitas, klasterisasi, k-means clustering, analisis data, keamanan.

CLUSTERING CRIME-PRONE AREAS IN INDONESIA USING THE K-MEANS METHOD

Abstract

Crime in Indonesia is a significant problem that affects various aspects such as security, social, and economic. However, mitigation efforts are often hampered by a lack of structured information about crime-prone areas. This study aims to overcome this problem by grouping provinces in Indonesia based on crime patterns using the K-Means Clustering algorithm. Crime statistics data for 2014-2023 from the Central Statistics Agency (BPS) were analyzed by determining the optimal number of clusters that resulted in five clusters with evaluation using Python and the Scikit-learn library. The results showed a Silhouette Score of 0.593, which reflects the formation of a fairly good cluster. This clustering provides data-driven guidance for the government in developing more targeted security policies to reduce crime rates in Indonesia.

Keywords: criminality, clustering, k-means clustering, data analysis, security

1. PENDAHULUAN

Kriminalitas berasal dari istilah "kriminal" yang merujuk pada serangkaian tindakan atau perilaku yang terkait dengan pelanggaran hukum atau kejahatan yang dapat dikenai sanksi pidana menurut peraturan undang-undang. Istilah kriminalitas merujuk pada tindakan yang melanggar hukum pidana atau dianggap sebagai kejahatan. Dalam hal ini, kriminalitas mencakup berbagai tindakan yang dianggap sebagai pelanggaran hukum atau perbuatan kriminal [1],[2]. Kejahatan seperti pembunuhan, penyerangan, pemerkosaan, dan lainnya semakin meningkat saat ini. Semakin meningkat angka kejahatan, semakin terancam rasa aman manusia. Indonesia menempati peringkat 70 sebagai negara dengan tingkat kriminalitas tinggi di dunia menurut

Numbeo 2022, sementara di Asia, Indonesia berada di peringkat 15. Berada di peringkat 5 di antara negara-negara di Asia Tenggara [3]. Di Indonesia, tingkat kejahatan berbeda-beda di setiap daerah, dipengaruhi oleh hal-hal seperti jumlah penduduk, kondisi ekonomi, tingkat pendidikan, dan situasi sosial [4]. Tingginya tingkat kejahatan seperti pembunuhan dan penipuan berdampak negatif yang signifikan. Kejahatan dapat menyebabkan kemiskinan dan mengganggu pertumbuhan ekonomi serta mengurangi rasa aman masyarakat [5].

Maka, penting untuk menganalisis dan mengelompokkan daerah-daerah yang rentan terhadap kejahatan guna mendukung penetapan kebijakan pencegahan dan penanganan yang lebih efektif. Penelitian ini berkontribusi dengan

mengidentifikasi pola kriminalitas di Indonesia secara sistematis, memberikan peta kerawanan yang dapat mendukung pengambilan kebijakan berbasis data. Berbeda dengan penelitian sebelumnya, studi ini memanfaatkan algoritma K-Means Clustering untuk memberikan analisis yang lebih komprehensif termasuk evaluasi menggunakan Silhouette Score, yang memberikan gambaran akurasi pengelompokan. Hasil penelitian diharapkan membantu pemerintah dan lembaga terkait dalam menyusun strategi keamanan yang lebih efektif sesuai dengan kebutuhan setiap wilayah.

Data Mining mengacu pada proses mengungkap informasi tersembunyi dalam basis data untuk menemukan wawasan baru dan pola yang bermakna dalam data. Disiplin ini menggunakan berbagai metodologi yang disesuaikan dengan tujuan khusus yang terkait dengan pemanfaatan kumpulan data, yang meliputi estimasi, prediksi, klasifikasi, pengelompokan, dan asosiasi. [6].

Clustering adalah teknik yang digunakan untuk mengkategorikan atau mengelompokkan sekumpulan objek berdasarkan atribut atau karakteristik yang sama dalam kaitannya dengan data lainnya. Ini adalah metode mendasar dalam bidang penambangan data, yang dicirikan oleh sifatnya yang tidak diawasi, yang menunjukkan bahwa pendekatan ini tidak memerlukan pelatihan sebelumnya. [7].

Algoritma K-Means Clustering adalah metode yang digunakan untuk mempartisi data ke dalam kluster-kluster berbeda berdasarkan kesamaan titik-titik data. Titik-titik data yang menunjukkan karakteristik serupa ditetapkan ke dalam kluster yang sama, sehingga memudahkan pengorganisasian dan analisis kumpulan data yang kompleks [8]. Metode ini sering digunakan untuk data besar yang proses pengelompokannya lebih cepat dibandingkan dengan metode hierarki [9]. Dengan algoritma ini, dilakukan pengelompokan berbagai wilayah di Indonesia berdasarkan tingkat kriminalitas. Ini tidak hanya membantu masyarakat untuk tetap waspada terhadap kriminalitas atau kejahatan, tetapi juga dapat mengevaluasi pemerintah dan memengaruhi kebijakan yang akan diterapkan oleh pemerintah [10].

Penelitian ini menggunakan algoritma K-Means Clustering untuk mengkategorikan wilayah-wilayah di Indonesia yang menunjukkan kerentanan terhadap aktivitas kriminal. Pengelompokan ini akan membantu pemerintah dan pihak keamanan dalam merancang strategi penanganan kriminalitas berdasarkan tingkat kerawanan di setiap daerah [11]. Analisis pola kejahatan melalui pengelompokan sistematis menawarkan wawasan berharga bagi lembaga pemerintah dan pasukan keamanan, yang memungkinkan mereka merancang tanggapan strategis yang disesuaikan dengan tingkat kerentanan kejahatan tertentu yang ada di wilayah geografis yang berbeda [12].

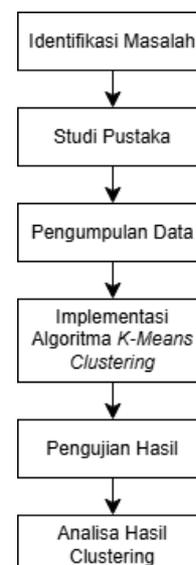
Penelitian terkait dilakukan oleh Mardhani Dwi Novianto dan Sri Andayani dalam penelitiannya yang mengelompokkan provinsi berdasarkan indikator kriminalitas nasional dengan metode SOM. Pengukuran kualitas kluster dilakukan menggunakan Davies-Bouldin Index untuk menilai seberapa baik kluster yang dihasilkan [13].

Sedangkan Jumadil, dkk mencoba mengidentifikasi daerah-daerah rawan kriminalitas di Sulawesi Tenggara menggunakan metode K-Means Clustering. Metode ini merupakan teknik data mining tanpa pengawasan (*unsupervised*) yang mengelompokkan data berdasarkan kesamaan karakteristiknya. Data yang digunakan diperoleh dari Badan Pusat Statistik (BPS) Sulawesi Tenggara, mencakup jumlah kejahatan dan persentase korban kejahatan di berbagai wilayah [1].

Penelitian ini memanfaatkan data kriminalitas terbaru dan mempertimbangkan berbagai faktor yang memengaruhi kriminalitas di setiap wilayah [12]. Diharapkan hasil pengelompokan ini bisa memberikan rekomendasi kepada pemerintah agar lebih mudah menentukan prioritas dan kebijakan dalam penanganan kriminalitas yang efektif dan efisien.

2. METODE PENELITIAN

Penelitian ini menggunakan metodologi kuantitatif dengan pendekatan data mining untuk meneliti provinsi-provinsi di Indonesia yang rawan terhadap tindak pidana, dengan menggunakan algoritma K-Means Clustering untuk analisis. Tahapan penelitian digambarkan pada Gambar 1.



Gambar 1. Kerangka Kerja Penelitian

2.1 Identifikasi Masalah

Pada tahap awal penelitian, fokusnya adalah mengidentifikasi isu-isu yang relevan dan melakukan tinjauan menyeluruh terhadap literatur yang ada,

termasuk jurnal ilmiah, artikel, sumber daring, buku, dan publikasi ilmiah. Tahap ini penting untuk mengumpulkan informasi relevan yang berkaitan dengan Data Mining, Klasifikasi, Algoritma K-Means, dan topik lainnya. Tujuan merumuskan masalah penelitian adalah untuk memastikan bahwa penelitian tetap fokus dan tidak menyimpang dari parameter penyelidikan yang ditentukan [14].

2.2 Pengumpulan Data

Data pada penelitian ini diperoleh dari Publikasi Statistik Kriminal 10 tahun terakhir yaitu pada tahun 2014-2023 yang diterbitkan oleh Bada Pusat Statistik (BPS) guna mengumpulkan data data jumlah tindak kriminal berdasarkan jenis kejahatan di setiap Kepolisian Daerah (Polda).

Penggunaan data yang mencakup hampir satu dekade ini memberikan gambaran yang komprehensif tentang tren kriminalitas di berbagai wilayah Indonesia. Pengambilan data dari BPS memastikan keakuratan dan validitas informasi, mengingat BPS adalah lembaga resmi pemerintah yang bertugas mengelola dan menyajikan data statistik nasional. Dataset ini sangat relevan untuk analisis tren jangka panjang serta perencanaan strategis dalam upaya pencegahan dan penanggulangan kriminalitas.

2.3 Implementasi Metode K-Means Clustering

Pengelompokan K-Means merupakan algoritma pengelompokan yang membagi himpunan data ke dalam kelompok-kelompok berbeda, sehingga titik-titik data dalam setiap kelompok memperlihatkan karakteristik yang serupa, sementara secara bersamaan memperlihatkan perbedaan dari titik-titik data di kelompok lain. 1. Pada implementasinya nanti akan menggunakan Python. Python adalah bahasa skrip berorientasi objek yang banyak digunakan untuk pengembangan perangkat lunak dan kompatibel dengan berbagai sistem operasi. Dalam konteks kontemporer, Python telah muncul sebagai bahasa yang menonjol di bidang ilmu data dan analisis data. Tren ini dapat dikaitkan dengan dukungannya yang kuat terhadap berbagai pustaka yang memfasilitasi analisis data, pembelajaran mesin, praproses data, dan visualisasi data [16]. Scikit-learn adalah pustaka sumber terbuka untuk analisis data yang mencakup berbagai algoritme penambangan data. Pustaka ini mencakup metodologi untuk klasifikasi, regresi, dan pengelompokan, serta berbagai teknik lainnya [7]. Langkah-langkah algoritma K-Means Clustering [15]:

1. Pemilihan acak sejumlah k titik data digunakan sebagai pusat kluster.
2. Jarak antara data dan pusat kluster dihitung menggunakan jarak Euclidean. Untuk menghitung jarak semua data ke setiap pusat kluster, kita dapat menggunakan teori jarak Euclidean yang dirumuskan sebagai berikut:

$$d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \dots(2)$$

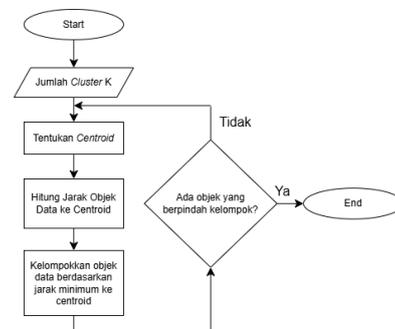
dimana, misalkan x_i adalah objek ke-i dalam himpunan objek x , y_i melambangkan objek ke-i di dalam himpunan objek y dan n melambangkan jumlah total objek yang sedang dipertimbangkan.

3. Titik-titik data ditetapkan ke kluster terdekat berdasarkan kedekatannya dengan titik pusat kluster tersebut.
4. Bila semua data telah ditetapkan ke kluster terdekat, maka akan ditentukan titik pusat kluster yang baru. Rumus perhitungan untuk titik pusat kluster yang baru adalah:

$$v = \sum_{i=1}^n x_i ; i = 1, 2, 3, \dots, n \dots\dots\dots(3)$$

Dimana, misalkan v melambangkan titik pusat kluster, X_i melambangkan objek ke- i , dan n melambangkan jumlah total objek yang termasuk dalam *cluster*.

5. Proses berulang dalam mengidentifikasi sentroid kluster dan menetapkan titik data ke klasternya masing-masing berlanjut hingga tidak ada lagi perubahan pada nilai sentroid. Diagram alir dari langkah langkah proses *K-Means Clustering* dapat dilihat pada gambar 2.



Gambar 2. Flowchart K-Means Clustering

2.4 Pengujian Hasil

Selama fase pengujian, dua metodologi digunakan: evaluasi manual terhadap hasil kluster dan pemanfaatan pemrograman Python. Selain itu, proses evaluasi menggabungkan Skor Silhouette, metrik statistik yang dirancang untuk menilai jumlah kluster optimal dalam kumpulan data tertentu [16]. *Skor Silhouette mengukur kohesi dan pemisahan kluster dengan menilai jarak setiap titik data dalam kluster relatif terhadap kluster lainnya. Metrik ini diperoleh dengan merata-ratakan jarak total, sehingga memberikan ukuran keseluruhan efektivitas pengelompokan* [17].

$$Silhouette\ Score = SI_k = \frac{1}{n} \sum_{i=1}^n \frac{(b_i - a_i)}{\max(a_i, b_i)} \dots\dots\dots(4)$$

2.5 Analisa Hasil Clustering

Analisis hasil dilakukan untuk menginterpretasikan temuan yang diperoleh dari data aktual. Hasil penilaian ini memberikan dasar bagi para pengambil keputusan, memfasilitasi pilihan yang tepat terkait kategorisasi area berdasarkan prevalensi aktivitas kriminal.

3. HASIL DAN PEMBAHASAN

3.1 Data Penelitian

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS). Dataset ini mencakup statistik kejahatan yang dikategorikan menurut provinsi untuk tahun 2014-2023, yang mencakup total 34 provinsi. Kriteria klasifikasi meliputi kejahatan terhadap nyawa, kejahatan terhadap integritas fisik, kejahatan terhadap moralitas, kejahatan terhadap harta benda, tindak pidana narkoba, serta penipuan dan korupsi. Berikut ini data jumlah tindakan kriminalitas dari tahun 2014-2023 untuk provinsi Aceh, Sumatera Utara, Sumatera Barat, Riau dan Jambi yang disajikan pada tabel 1-10.

Tabel 1. Tindakan Kriminalitas 2014

Provinsi	A1	A2	A3	A4	A5	A6	A7	A8
aceh	37	48	89	22	18	74	82	91
	6	0			2	8	0	5
sumatera utara	34	60	19	13	84	70	21	22
	05		0	0	3	14	06	39
sumatera barat	74	38	70	27	41	21	31	97
	3	7			5	39	6	3
riau	56	26	39	31	38	12	10	51
	7	5			4	59	37	4
jambi	29	84	41	2	20	79	23	27
	9				3	6	8	6

Tabel 2. Tindakan Kriminalitas 2015

Provinsi	A1	A2	A3	A4	A5	A6	A7	A8
aceh	21	18	18	62	16	33	68	11
		53	8		1	23	8	07
sumatera utara	94	63	24	17	12	16	27	45
		37	0	2	07	11	32	31
						9		
sumatera barat	12	28	35	69	43	82	30	22
		34	0		7	53	5	07
riau	30	15	27	15	45	40	94	14
		02	7		3	62	3	69
jambi	22	92	80	13	23	34	27	14
		4			4	00	0	29

Tabel 3. Tindakan Kriminalitas 2016

Provinsi	A1	A2	A3	A4	A5	A6	A7	A8
aceh	40	20	23	64	49	37	94	91
		75	5			28	8	1
sumatera utara	11	55	15	10	15	13	47	50
	8	82	4	5	91	78	11	23
						0		
sumatera barat	18	33	36	12	52	83	97	27
		46	9	7	0	53	4	42
riau	49	13	28	4	50	37	12	12
		86	4		3	47	94	67
jambi	23	11	95	23	38	45	37	20
		67			7	72	1	83

Tabel 4. Tindakan Kriminalitas 2017

Provinsi	A1	A2	A3	A4	A5	A6	A7	A8
aceh	43	2.0	29	60	19	4.0	14	1.3
		03	1		1	13	02	78
sumatera utara	16	5.7	18	21	1.1	13.	5.5	5.1
	1	02	6	2	60	52	91	33
						6		
sumatera barat	21	2.7	34	14	53	8.2	72	2.0
		33	0	8	2	11	4	87
riau	20	1.1	17	3	39	3.1	1.4	1.1
		99	3		1	29	51	17
jambi	27	1.1	11	82	37	3.8	46	1.7
		43	1		5	39	2	90

Tabel 5. Tindakan Kriminalitas 2018

Provinsi	A1	A2	A3	A4	A5	A6	A7	A8
aceh	27	1.7	31	38	21	3.2	1.5	1.3
		64	1		9	38	44	38
sumatera utara	92	5.6	38	19	90	12.	5.9	5.2
		33	4		8	36	07	97
						1		
sumatera barat	14	2.3	34	15	46	7.5	59	1.8
		71	3	8	3	11	2	81
riau	28	79	97	11	26	2.3	1.1	96
		1			3	89	73	6
jambi	24	1.0	11	22	27	3.9	71	1.8
		15	3	5	4	13	2	65

Tabel 6. Tindakan Kriminalitas 2019

Provinsi	A1	A2	A3	A4	A5	A6	A7	A8
aceh	22	1.4	26	41	15	2.9	1.6	1.4
		75	2		6	59	25	71
sumatera utara	10	5.2	32	88	70	10.	6.2	4.6
	7	40	1		2	81	62	8
						8		
sumatera barat	14	2.1	39	13	38	7.2	80	1.8
		53	2	2	7	30	6	47
riau	30	85	98	16	30	2.3	1.7	1.0
		2			1	33	17	60
jambi	23	57	90	53	11	2.2	43	95
		1			9	83	9	9

Tabel 7. Tindakan Kriminalitas 2020

Provinsi	A1	A2	A3	A4	A5	A6	A7	A8
aceh	26	12	26	28	14	25	15	11
		56	7		0	04	82	20
sumatera utara	10	48	21	10	68	86	62	49
	4	17	6	5	5	93	01	04
sumatera barat	16	19	33	10	27	62	80	13
		88	3	4	7	78	6	52
riau	19	91	11	8	25	19	16	98
		8	8		4	68	71	7
jambi	15	73	69	31	16	28	40	12
		9			0	70	7	06

Tabel 8. Tindakan Kriminalitas 2021

Provinsi	A1	A2	A3	A4	A5	A6	A7	A8
aceh	15	12	23	29	14	24	15	11
		06	2		5	08	96	30
sumatera utara	99	62	77	55	78	10	59	55
		07	4		0	91	32	62
						6		
sumatera barat	20	17	29	97	18	37	91	98
		24	4		5	32	3	9
riau	21	10	14	11	41	31	16	10
		13	3		3	52	46	95
jambi	11	53	69	9	87	17	61	76
		1				70	9	7

Tabel 9. Tindakan Kriminalitas 2022

Provinsi	A1	A2	A3	A4	A5	A6	A7	A8
aceh	29	10	20	23	10	23	12	10
		74	4		7	80	08	23
sumatera utara	96	42	90	86	54	12	59	53
		87	4		0	43	49	96
						3		
sumatera barat	10	12	34	48	98	19	99	73
		99	0			71	4	9
riau	24	72	41	5	25	26	16	94
		8			5	22	70	6
jambi	15	39	46	3	73	15	36	59
		0				45	1	7

Tabel 10. Tindakan Kriminalitas 2023

Provinsi	A1	A2	A3	A4	A5	A6	A7	A8
aceh	24	11	21	51	57	26	12	10
		78	7			59	07	12
sumatera utara	84	59	45	12	63	16	41	53
		34	3	8	6	34	62	76
						7		
sumatera barat	8	12	15	30	52	15	92	55
		98	7			38	2	1
riau	36	10	17	46	32	49	14	12
		23	6		2	23	70	14
jambi	15	58	67	27	76	16	62	46
		2				42	6	9

Berikut adalah keterangan kode kriminalitas yang ditandai dengan A1-18 dan deskripsi kejahatan dari setiap klasifikasi kejahatan yang disajikan pada tabel 11.

Tabel 11. Keterangan kode Kriminalitas dan Klasifikasi Kejahatan

Kode	Klasifikasi Kejahatan	Jenis Kejahatan
A1	Kejahatan terhadap nyawa	Pembunuhan
A2	Kejahatan terhadap Fisik/Badan	Penganiayaan Berat Penganiayaan Ringan Kekerasan dalam Rumah Tangga
A3	Kejahatan terhadap Kesusilaan	Perkosaan Pencabulan
A4	Kejahatan terhadap Kemerdekaan Orang	Penculikan Mempekerjakan Anak di Bawah Umur
A5	Kejahatan terhadap Hak Milik/Barang dengan Penggunaan Kekerasan	Pencurian dengan Kekerasan Pencurian dengan Kekerasan Menggunakan Senjata Api (Senpi) Pencurian dengan Kekerasan Menggunakan Senjata Tajam (Sajam)
A6	Kejahatan Terhadap Hak Milik/Barang	Pencurian Pencurian dengan pemberatan Pencurian Kendaraan Bermotor Pengrusakan/Penghancuran Barang Pembakaran dengan Sengaja Penadahan
A7	Kejahatan Terkait Narkotika	Narkotika dan Psicotropika
A8	Kejahatan terkait Penipuan,	Penipuan/Perbuatan Curang Penggelapan

Penggelapan, dan Korupsi	Korupsi
--------------------------	---------

3.2 Jumlah Cluster (K)

Dalam penelitian ini, jumlah kluster yang ditetapkan adalah 2. Diasumsikan akan terbentuk dua kelompok yang memiliki karakteristik sangat berbeda satu sama lain dalam tingkatan kriminalitas dimana Cluster 1 adalah daerah dengan tingkat kriminalitas tinggi dan Cluster 2 adalah daerah dengan tingkat kriminalitas rendah.

Pemilihan ini didukung oleh metode evaluasi Elbow, yang menunjukkan bahwa k=2 memberikan hasil optimal dengan perbedaan yang jelas antar kelompok. Selain itu, dua kluster memudahkan interpretasi data dan penerapan kebijakan yang lebih sederhana, seperti fokus pengamanan intensif untuk wilayah berisiko tinggi dan pencegahan di wilayah dengan risiko rendah.

3.3 Centroid Data

Centroid data diambil secara acak untuk Pusat cluser 1 (C₁) diambil dari provinsi Jawa Barat dan cluser 2 (C₂) diambil dari provinsi Banten.

Tabel 12. Hasil Perhitungan Jarak Data

Provinsi	DC1	DC2
aceh	15438,28149	4779,340213
sumatera utara	22460,33429	35686,61043
sumatera barat	11873,94896	11290,93447
riau	14071,09971	6009,403573
jambi	15849,74987	3674,261577
sumatera selatan	10179,94287	11969,30166
bengkulu	19442,31504	2251,482579
lampung	16684,34085	5642,433533
kep. Bangka belitung	20991,93216	3844,34267
kepulauan riau	18948,49087	1751,664082
metro jaya	15152,04082	25715,64757
jawa barat	0	17908,39133
jawa tengah	11678,44509	8458,87229
di yogyakarta	17374,65733	2366,897152
jawa timur	14966,62629	18983,68885
banten	17908,39133	0
bali	19323,21042	2457,233252
nusa tenggara barat	18957,46665	2370,682873
nusa tenggara timur	18554,40764	4621,869704
kalimantan barat	17388,02107	1907,512921
kalimantan tengah	20334,60217	3279,559418
kalimantan selatan	19201,08198	3348,713811
kalimantan timur	16932,73889	3727,176407
kalimantan utara	22056,17569	4814,363543
sulawesi utara	18616,71316	6052,548487
sulawesi tengah	16098,79991	3876,378691
sulawesi selatan	11468,5924	14838,81828
sulawesi tenggara	19714,1334	3364,949911
gorontalo	20276,47015	3820,314888
sulawesi barat	21549,11514	4439,411352
maluku	21042,70026	4185,123145
maluku utara	21697,71942	4817,780053
papua	16416,39013	5233,956794
papua barat	20711,44107	3953,471599

3.4 Menghitung Jarak Terdekat Centroid

Rumus jarak Euclidean digunakan untuk menghitung jarak suatu objek ke pusat setiap centroid

guna menemukan jarak terdekat ke centroid setiap data. Setelah menyelesaikan perhitungan jarak antara titik data dan pusat kluster, seperti yang diuraikan dalam metodologi sebelumnya, nilai jarak yang dihasilkan disajikan dalam Tabel 12.

3.5 Mengelompokkan Data Kedalam Cluster

Setelah selesai melakukan perhitungan jarak dari setiap provinsi, maka selanjutnya dikelompokkan kedalam cluster yang disajikan pada tabel 13.

Tabel 13. Pengelompokan Data ke Centroid

Provinsi	Cluster
aceh	C2
sumatera utara	C1
sumatera barat	C2
riau	C2
jambi	C2
sumatera selatan	C1
bengkulu	C2
lampung	C2
kep. Bangka belitung	C2
kepulauan riau	C2
metro jaya	C1
jawa barat	C1
jawa tengah	C2
di yogyakarta	C2
jawa timur	C1
banten	C2
bali	C2
nusa tenggara barat	C2
nusa tenggara timur	C2
kalimantan barat	C2
kalimantan tengah	C2
kalimantan selatan	C2
kalimantan timur	C2
kalimantan utara	C2
sulawesi utara	C2
sulawesi tengah	C2
sulawesi selatan	C1
sulawesi tenggara	C2
gorontalo	C2
sulawesi barat	C2
maluku	C2
maluku utara	C2
papua	C2
papua barat	C2

Hasil klasterisasi dengan metode K-Means Clustering pada data daerah kriminalitas di Indonesia memberikan pengelompokan provinsi berdasarkan pola atau karakteristik tertentu terkait dengan tingkat kriminalitas.

Hasil klasterisasi menunjukkan pengelompokan provinsi di Indonesia berdasarkan karakteristik kriminalitas yang serupa. Cluster 1 mencakup provinsi seperti Sumatera Utara, Jawa Barat, dan Jawa Timur, yang kemungkinan memiliki tingkat kriminalitas lebih tinggi atau pola kejahatan khas urbanisasi dan ekonomi. Sementara itu, Cluster 2 mencakup sebagian besar provinsi lain, seperti Aceh, Bali, dan Papua, yang cenderung memiliki tingkat kriminalitas lebih rendah atau pola kejahatan terkait wilayah pedesaan dan tradisional. Informasi ini membantu memahami kesamaan antar provinsi

dalam satu kluster, seperti tingkat urbanisasi, kepadatan penduduk, atau jenis kejahatan dominan. Data ini bermanfaat untuk merancang kebijakan keamanan yang lebih spesifik dan efisien berdasarkan kebutuhan masing-masing wilayah.

3.6 Pengujian Hasil

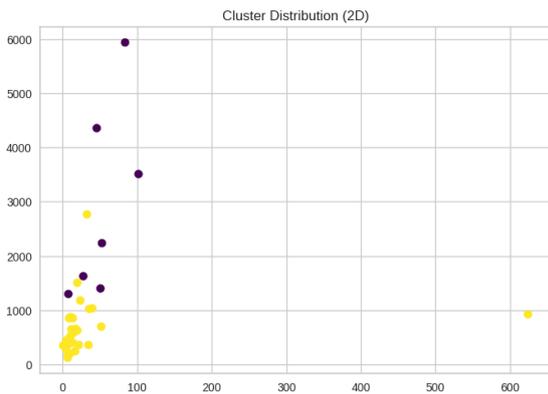
Berikut adalah hasil yang didapatkan dari analisis klasterisasi menggunakan Python yang disajikan pada tabel 14.

Tabel 14. Pengelompokan Cluster pada analisis menggunakan Python

Provinsi	Cluster
aceh	C2
sumatera utara	C1
sumatera barat	C1
riau	C2
jambi	C2
sumatera selatan	C1
bengkulu	C2
lampung	C2
kep. Bangka belitung	C2
kepulauan riau	C2
metro jaya	C1
jawa barat	C1
jawa tengah	C2
di yogyakarta	C2
jawa timur	C1
banten	C2
bali	C2
nusa tenggara barat	C2
nusa tenggara timur	C2
kalimantan barat	C2
kalimantan tengah	C2
kalimantan selatan	C2
kalimantan timur	C2
kalimantan utara	C2
sulawesi utara	C2
sulawesi tengah	C2
sulawesi selatan	C1
sulawesi tenggara	C2
gorontalo	C2
sulawesi barat	C2
maluku	C2
maluku utara	C2
papua	C2
papua barat	C2

Hasil dari pengelompokan ini menunjukkan bahwa mayoritas provinsi dikelompokkan dengan benar sesuai dengan pengelompokan manual, namun ada perbedaan dalam pengelompokan provinsi Sumatera Barat yang masuk dalam cluster C1 di Python, sedangkan dalam pengelompokan manual dikelompokkan dalam C2. Perbedaan ini bisa disebabkan oleh perbedaan dalam metode atau parameter yang digunakan dalam algoritma *clustering* Python, serta perbedaan cara centroid dihitung dan dipilih.

Visualisasi distribusi cluster dapat dilihat pada gambar 3. Titik berwarna kuning merupakan titik-titik data yang masuk ke dalam cluster C1 dan titik berwarna ungu merupakan titik-titik data yang masuk ke dalam cluster C2.



Gambar 3. Visualisasi Distribusi Cluster

Dalam pengujian hasil dilakukan pengujian Silhouette Score untuk mengevaluasi kualitas clustering. Hasilnya didapatkan Silhouette Score tertinggi untuk jumlah cluster terbaik adalah pada $k = 2$, dengan perolehan score 0,593. Hasil tersebut menunjukkan bahwa hasil clustering cukup baik. Sebagian besar objek dalam cluster berada cukup dekat dengan cluster mereka sendiri dan terpisah dari cluster lainnya.

3.7 Analisa Hasil Clustering

Dari data di atas didapatkan hasil clustering yaitu provinsi yang masuk ke dalam Cluster 1 yaitu Sumatera Utara, Sumatera Barat, Sumatera Selatan, Metro Jaya (Jakarta), Jawa Barat dan Sulawesi Selatan. Provinsi-provinsi pada cluster ini memiliki karakteristik yaitu dengan tingkat kejahatan cenderung jauh lebih tinggi dibandingkan Cluster C2. Kejahatan terhadap nyawa, fisik, dan penipuan memiliki angka rata-rata yang signifikan, dengan dominasi pada kejahatan ekonomi seperti penipuan dan barang tanpa kekerasan. Hal ini mencerminkan kompleksitas ekonomi di daerah urban yang memberikan lebih banyak peluang bagi tindak kejahatan. Selain itu, kejahatan narkoba juga menjadi salah satu isu besar, mencerminkan jaringan distribusi narkoba yang lebih kuat di kawasan ini. Tantangan sosial khas wilayah perkotaan, seperti ketimpangan sosial, kemiskinan, dan tekanan kehidupan kota, turut berkontribusi pada tingginya tingkat kriminalitas di cluster ini.

Sedangkan, sisanya merupakan bagian dari Cluster 2 di antaranya Aceh, Riau, Jambi, dan 24 provinsi lainnya. Provinsi-provinsi pada cluster ini memiliki karakteristik di antaranya dengan tingkat kejahatan lebih rendah, mencerminkan kondisi sosial yang lebih stabil dengan tekanan kriminalitas yang lebih ringan. Kejahatan barang tanpa kekerasan menjadi kategori yang paling dominan, tetapi jumlahnya masih jauh lebih rendah dibandingkan Cluster C1. Interaksi sosial yang lebih erat, pengawasan komunitas yang kuat, dan aktivitas ekonomi yang lebih sederhana menjadi faktor penting dalam menekan angka kejahatan. Selain itu, distribusi

narkoba di wilayah Cluster C2 juga lebih terbatas, sehingga kasus-kasus terkait narkoba jarang ditemukan dibandingkan dengan Cluster C1. Secara keseluruhan, Cluster C2 mencerminkan wilayah dengan dinamika sosial yang lebih harmonis dan tantangan kriminalitas yang lebih ringan.

4. KESIMPULAN

Metode *K-Means Clustering* berhasil diterapkan untuk mengelompokkan pola kriminalitas di Indonesia dengan mengelompokkan provinsi-provinsi ke dalam 2 cluster berdasarkan kesamaan karakteristik dan tingkat kejahatan. Dengan perolehan *Silhouette Score* yaitu 0.593 yang menunjukkan bahwa cluster sudah cukup baik.

Hasil penelitian ini dapat menjadi dasar bagi pihak kepolisian, pemerintah dan lainnya yang berwenang dalam pengambilan kebijakan agar lebih terfokus dan efisien yang disesuaikan dengan karakteristik setiap cluster.

DAFTAR PUSTAKA

- [1] Muh. A. Z. R. Ramadhan, R. A. Saputra, and J. Nangi, "Klasterisasi Daerah Rawan Kriminalitas di Sulawesi Tenggara Menggunakan Metode K-Means Clustering," *JOINTER - Journal of Informatics Engineering*, vol. 05, no. 01, pp. 1–6, 2024.
- [2] L. Suriani, "Pengelompokan Data Kriminal Pada Poldasu Menentukan Pola Daerah Rawan Tindak Kriminal Menggunakan Data Mining Algoritma K-Means Clustering," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 1, no. 2, p. 151, Jan. 2020, doi: 10.30865/json.v1i2.1955.
- [3] J. Inayah, D. A. S. N. Maghfiroh, and D. C. R. Novitasari, "Clustering Daerah Rawan Kriminalitas Menggunakan Algoritma Fuzzy C-Means," *Jurnal Ilmiah Informatika Komputer*, vol. 27, no. 2, pp. 95–106, 2022, doi: 10.35760/ik.2022.v27i2.6019.
- [4] A. A. Dunda, A. Purwoto, A. Choirunnisa, R. A. Falah, and R. Y. Putri, "Faktor-Faktor yang Mempengaruhi Jumlah Tindak Kriminal di Sumatera Utara Tahun 2022," *Madani: Jurnal Ilmiah Multidisipliner*, vol. 2, no. 3, pp. 543–549, 2024, doi: 10.5281/zenodo.12171133.
- [5] T. Palokoto, E. Y. Purwanti, and Y. B. Mudakir, "Analisis Konsekuensi Ekonomi Akibat Kejahatan di Indonesia," *Jurnal Ekonomi dan Pembangunan Indonesia*, vol. 20, no. 2, pp. 146–159, Jul. 2020, doi: 10.21002/jepi.2020.09.
- [6] P. M. S. Tarigan, J. T. Hardinata, H. Qurniawan, M. Safii, and R. Winanjaya, "Implementasi Data Mining Menggunakan Algoritma Apriori dalam Menentukan Persediaan Barang (Studi Kasus : Toko Sinar Harahap)," *Just IT: Jurnal Sistem Informasi, Teknologi Informasi dan Komputer*, vol. 12, no. 2, pp. 51–61, 2022, [Online]. Available: <https://jurnal.umj.ac.id/index.php/just-it/index>
- [7] D. A. Manalu and G. Gunadi, "Implementasi Metode Data Mining K-Means Clustering terhadap Data Pembayaran Transaksi

- Menggunakan Bahasa Pemrograman Python pada CV Digital Dimensi,” *Infotech: Journal of Technology Information*, vol. 8, no. 1, pp. 43–54, Jun. 2022, doi: 10.37365/jti.v8i1.131.
- [8] L. Suriani, “Pengelompokan Data Kriminal Pada Poldasu Menentukan Pola Daerah Rawan Tindak Kriminal Menggunakan Data Mining Algoritma K-Means Clustering,” *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 1, no. 2, p. 151, Jan. 2020, doi: 10.30865/json.v1i2.1955.
- [9] P. S. Rosiana, A. A. Mohsa, M. A. Fadila, and J. H. Jaman, “Visualisasi Data Kejahatan di Kabupaten Karawang Berdasarkan Jenis Kriminalitas dengan Algoritma Clustering K-Means,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 11, no. 3, pp. 2830–7062, 2023, doi: 10.23960/jitet.v11i3%20s1.3347.
- [10] M. D. Novianto and S. Andayani, “Analisis Cluster dengan Metode SOM untuk Pengelompokan Provinsi di Indonesia berdasarkan Indikator Kriminalitas,” *Jurnal Kajian dan Terapan Matematika*, vol. 9, no. 1, pp. 39–53, 2023, [Online]. Available: <https://www.numbeo.com/>
- [11] Statistik Kriminal oleh Badan Pusat Statistik 2023. 2023.
- [12] R. M. Salsabila and A. Syarief, “Analisis Spasial Tingkat Kerawanan Kriminalitas Menggunakan Metode Nearest Neighbor Analysis Dan Kernel Density Di Kota Pekanbaru,” *Jurnal Pendidikan Tambusai*, vol. 8, no. 2, pp. 28844–28850, 2024.
- [13] M. D. Novianto and S. Andayani, “Analisis Cluster dengan Metode SOM untuk Pengelompokan Provinsi di Indonesia berdasarkan Indikator Kriminalitas,” *Jurnal Kajian dan Terapan Matematika*, vol. 9, no. 1, pp. 39–53, 2023, [Online]. Available: <https://www.numbeo.com/>
- [14] T. Hidayat, “Klasifikasi Data Jamaah Umroh Menggunakan Metode K-Means Clustering,” *Jurnal Sistim Informasi dan Teknologi*, vol. 4, no. 1, pp. 19–24, Feb. 2022, doi: 10.37034/jsisfotek.v4i1.115.
- [15] H. Kurniawan, S. Defit, and Sumijan, “Data Mining Menggunakan Metode K-Means Clustering Untuk Menentukan Besaran Uang Kuliah Tunggal,” *Journal of Applied Computer Science and Technology (JACOST)*, vol. 1, no. 2, pp. 80–89, Dec. 2020, doi: 10.52158/jacost.v1i2.102.
- [16] D. A. S. Simamora, M. T. Furqon, and B. Priyambadha, “Clustering Data Kejadian Tsunami Yang Disebabkan Oleh Gempa Bumi Dengan Menggunakan Algoritma K-Medoids,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 1, no. 8, pp. 635–640, 2017, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [17] A. H. Yusup and W. Maharani, “Model Prediksi Kepribadian Berdasarkan Tweet dan Kategori Big Five dengan Metode Agglomerative Hierarchical Clustering,” *TELKATIKA*, vol. 1, no. 1, p. 2021, 2021.