

## Pengelompokkan Tindakan Kriminalitas di Indonesia dengan K-Medoids Menggunakan Algoritma Artificial Bee Colony

Adinda Pratiwi Musa<sup>1\*</sup>, Novianita Achmad<sup>2</sup>, Salmun K. Nasib<sup>3</sup>

<sup>1,3</sup> Program Studi Statistika, Universitas Negeri Gorontalo, Indonesia

<sup>2</sup> Program Studi Matematika, Universitas Negeri Gorontalo, Indonesia

<sup>1</sup>adindayputri11@gmail.com\*, <sup>2</sup>novianita.achmad@ung.ac.id, <sup>3</sup>salmun@ung.ac.id

Diajukan: 26 Desember 2025 | Direvisi: 9 Januari 2026 | Diterima: 12 Januari 2026 | Diterbitkan: 9 Februari 2026

### Abstract

*This research aims to classify regional police forces in Indonesia based on the level and characteristics of criminal activity by applying the K-Medoids method optimized using the Artificial Bee Colony (ABC) algorithm. This grouping is intended to identify the dominant crime patterns in each region and evaluate the effectiveness of the methods used in producing representative clusters. The analysis results show that the K-Medoids-ABC method produces three main clusters, with the distribution of each consisting of 7 regional police departments in cluster 1, 5 regional police departments in cluster 2, and 21 regional police departments in cluster 3. Cluster validation using the Silhouette Index (SI) yielded a value of 0.387, indicating that the clustering results fall into the weak structure category, meaning the cluster structure is formed but with weak separation (Weak Separation). Cluster 1 shows a moderate and relatively even crime rate, Cluster 2 is dominated by crimes against life and crimes of fraud, embezzlement, and corruption, while Cluster 3 shows low values across all variables, with the lowest values for violent property crimes and drug-related crimes. This cluster reflects regions with relatively safe conditions, as evidenced by very low crime rates. These differences in characteristics between clusters reflect the diversity of factors causing crime in each region and have important implications for formulating more contextual and targeted crime prevention strategies.*

**Keywords:** Artificial Bee Colony, Clustering, Criminal Activity, K-Medoids, Regional Police Department, Silhouette Index.

### Abstrak

Penelitian ini bertujuan untuk mengelompokkan kepolisian daerah di Indonesia berdasarkan tingkat dan karakteristik tindak kriminalitas dengan menerapkan metode K-Medoids yang dioptimasi menggunakan algoritma Artificial Bee Colony (ABC). Pengelompokan ini dimaksudkan untuk mengidentifikasi pola-pola kriminalitas yang dominan di masing-masing wilayah serta mengevaluasi efektivitas metode yang digunakan dalam menghasilkan kluster yang representatif. Hasil analisis menunjukkan bahwa metode K-Medoids-ABC menghasilkan tiga kluster utama, dengan distribusi masing-masing terdiri dari 7 kepolisian daerah pada kluster 1, 5 kepolisian daerah pada kluster 2, dan 21 kepolisian daerah pada kluster 3. Validasi kluster menggunakan Silhouette Index (SI) menghasilkan nilai sebesar 0,387, yang mengindikasikan hasil pengelompokan termasuk kategori struktur lemah, yang artinya struktur kluster terbentuk namun dengan tingkat pemisahan yang lemah (*Weak Separation*). Kluster 1 menunjukkan tingkat kriminalitas yang sedang dan relatif merata, kluster 2 didominasi oleh kejahatan terhadap nyawa dan kejahatan terhadap penipuan, penggelapan, korupsi, sedangkan kluster 3 menunjukkan nilai rendah di seluruh variabel, nilai yang paling rendah ditunjukkan oleh kejahatan terhadap hak milik dengan penggunaan kekerasan dan kejahatan terkait narkoba. Kluster ini mencerminkan wilayah dengan kondisi relatif aman dilihat dari tingkat kejahatan yang sangat rendah. Perbedaan karakteristik antar kluster ini mencerminkan keberagaman faktor penyebab kriminalitas di tiap daerah dan memberikan implikasi penting bagi perumusan strategi penanggulangan kejahatan yang lebih kontekstual dan tepat sasaran.

**Kata kunci:** Artificial Bee Colony, Clustering, Kepolisian daerah, K-Medoids, Silhouette Index, Tindak kriminalitas.

*This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0). Copyright (C) Author's.*



## 1. PENDAHULUAN

*Clustering* dapat dibedakan menjadi dua jenis, yaitu *hard clustering* dan *soft clustering*. *Hard Clustering* berfokus pada memaksimalkan kemiripan antar objek yang diamati sehingga membentuk satu kluster tertentu. Sebaliknya, *soft clustering* (atau *fuzzy clustering*) mengadopsi prinsip logika *fuzzy*, di mana suatu objek dapat termasuk dalam lebih dari satu kluster dengan tingkat keanggotaan tertentu. Untuk menjawab tantangan *clustering* pada data berdimensi tinggi dan volume besar, pendekatan *fuzzy clustering* menjadi Solusi yang relevan [1].

Menurut [2] *clustering* merupakan teknik pengelompokan data ke dalam beberapa kelompok berdasarkan kriteria kesamaan yang telah ditetapkan. Klaster adalah kumpulan objek data yang memiliki kemiripan tinggi di dalam satu kelompok, tetapi berbeda secara signifikan dengan objek pada klaster lain. Pengelompokan dilakukan dengan tujuan memaksimalkan kesamaan antarobjek dalam klaster yang sama sekaligus memaksimalkan perbedaan antar-klaster. Ukuran kesamaan biasanya diperoleh dari nilai atribut yang mendeskripsikan objek, sehingga data direpresentasikan sebagai titik dalam ruang multidimensi.

Analisis klaster merujuk pada sekumpulan metode untuk mengelompokkan objek atau data berdasarkan kesamaan dan karakteristiknya [3]. Tujuannya adalah mengidentifikasi kelompok objek dengan ciri serupa yang dapat dipisahkan dari klaster lain, sehingga tidak terjadi tumpang tindih atau penggabungan yang tidak diinginkan antarobjek [4]. Analisis klaster terbagi menjadi dua pendekatan utama yaitu, metode hierarkis (*hierarchical clustering*) dan non-hierarkis (*non-hierarchical clustering*). Metode hierarkis menggabungkan objek-objek yang sangat mirip secara bertahap hingga membentuk struktur berbentuk pohon (*dendrogram*), sedangkan metode non-hierarkis menetapkan jumlah klaster terlebih dahulu, kemudian menempatkan objek ke dalam klaster yang sesuai [5]. Metode non-hierarkis mencakup dua algoritma utama, yaitu *K-means* dan *Fuzzy C-means*. Di antara keduanya, *K-means* paling sering digunakan, tetapi memiliki kelemahan, terutama dalam menangani outlier atau data yang tidak memiliki keterkaitan kuat dengan klaster lain. Untuk mengatasi keterbatasan tersebut, Kaufman dan Rousseeuw memperkenalkan algoritma *K-Medoids* pada tahun 1987.

*K-Medoids* merupakan variasi dari *K-Means* yang menggunakan titik data nyata sebagai pusat klaster, sehingga lebih *robust* terhadap outlier dibandingkan *K-Means* yang bergantung pada nilai rata-rata. Algoritma ini mengelompokkan data ke dalam sejumlah klaster dengan meminimalkan total jarak antarobjek dalam satu klaster. Meski demikian, *K-Medoids* cenderung lebih mahal secara komputasi karena mengevaluasi setiap titik data sebagai kandidat pusat. Validasi hasil pengelompokan menjadi langkah krusial untuk memastikan klaster yang dihasilkan mencerminkan struktur data secara akurat, sehingga meningkatkan keandalan analisis [6].

*Algoritma Artificial Bee Colony (ABC)*, yang diperkenalkan oleh Karaboga pada tahun 2005, adalah teknik optimasi yang meniru pola pencarian makanan lebah madu. Dalam ABC, koloni lebah terdiri atas tiga peran utama yaitu, lebah pekerja, lebah pengamat, dan lebah pengintai, yang berkolaborasi untuk mengeksplorasi dan mengeksploitasi ruang pencarian demi mencapai solusi optimal. Keunggulan ABC meliputi kesederhanaan implementasi, fleksibilitas tinggi, serta kemampuan menangani masalah non-linear dan ruang solusi berdimensi besar. Kemampuan pencarian paralelnya juga membuatnya efisien untuk berbagai aplikasi, termasuk pengelompokan data seperti analisis pola curah hujan. Penelitian oleh [7] menunjukkan bahwa ABC mengungguli algoritma *Particle Swarm Optimization (PSO)* dalam menemukan optimum global, baik dari segi efisiensi komputasi maupun tingkat akurasi solusi.

## 2. METODE PENELITIAN/ALGORITMA

*Principal Component Analysis (PCA)* adalah teknik yang dirancang untuk menangani multikolinearitas tanpa harus menghapus variabel independen yang saling berkorelasi [8]. PCA bertujuan menyederhanakan data melalui reduksi dimensi dengan mengubah variabel asli menjadi komponen-komponen baru yang ortogonal (tidak berkorelasi) [9]. Jumlah komponen utama yang dipertahankan ditentukan berdasarkan nilai *eigenvalue* di atas 1 atau proporsi varians kumulatif minimal 80% menurut Human dalam [10]. Penerapan PCA sebelum dilakukan analisis cluster bertujuan untuk memastikan pembentukan klaster dapat mencerminkan pola perbedaan karakteristik antar kelompok secara objektif, tanpa dipengaruhi oleh redundansi informasi akibat multikolinearitas. Adapun asumsi yang harus dipenuhi oleh, yaitu kecukupan sampel yang divalidasi melalui uji *Kaiser-Meyer-Olkin (KMO)* serta adanya korelasi signifikan antarvariabel independen yang diuji melalui *Bartlett's Test of Sphericity* [11].

Uji KMO digunakan untuk mengukur kecukupan data dalam rangka penerapan analisis PCA [11].  
Hipotesis :

H<sub>0</sub> = Sampel tidak layak untuk dianalisis lebih lanjut

H<sub>1</sub> = Sampel layak untuk dianalisis lebih lanjut

Statistik Uji :

$$KMO = \frac{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2}{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2 + \sum_{i=1}^p \sum_{j=1}^p p_{ij}^2} \quad (1)$$

Keterangan

$r_{ij}$  : Koefisien korelasi sederhana antara variabel ke-i dan ke-j

$p_{ij}$  : Koefisien korelasi parsial antara variabel ke-i dan ke-j

Keputusan

$H_0$  : diterima jika  $KMO < 0,5$

$H_0$  : ditolak jika  $KMO > 0,5$

**Tabel 1.** Kriteria Keputusan

Nilai KMO	Interpretasi
0.91 - 1.00	sangat layak
0.81 - 0.90	layak
0.61 - 0.80	lebih dari cukup
0.51 - 0.60	cukup
0.00 - 0.50	tidak layak

Untuk melihat apakah matriks korelasinya merupakan matriks identitas yaitu dengan menggunakan Uji *Bartlett Sphericity* [12].

$$X_{obs}^2 = - \left[ (N - 1) - \frac{(2p+5)}{6} \right] \ln|R| \quad (2)$$

derajat Kebebasan (*degree of freedom*):

dimana:

$N$  : Total Observasi

$p$  : Banyak variabel

$|R|$  : Determinan matriks korelasi

Keputusan :

$H_0$  : diterima jika  $X^2 < X_{tabel}(p - 1)/2$

$H_0$  : ditolak jika  $X^2 \geq X_{tabel}(p - 1)/2$  atau  $\text{sig} < \alpha$

Metode *Elbow* merupakan salah satu pendekatan untuk menentukan jumlah kluster optimal, dengan prinsip mengidentifikasi titik “siku” pada kurva persentase varians yang dijelaskan oleh jumlah kluster. Pada penelitian ini menggunakan data nilai *Principal Component* untuk penentuan jumlah cluster menggunakan metode *elbow*. Titik siku tercapai ketika penurunan nilai *Sum of Squared Error* (SSE) dari kluster ke-1 ke kluster ke-2 memberikan sudut dalam grafik dengan nilai yang mengalami penurunan terbesar maka nilai tersebut merupakan cluster nilai terbaik [13]. Nilai SSE untuk setiap jumlah kluster dihitung menggunakan rumus yang dikemukakan oleh Christopher Bishop (2006) dalam [14] :

$$SEE = \sum_{k=1}^K \sum_{i \in X_k} |X_i - C_k|^2 \quad (3)$$

dimana:

$K$  : Cluster ke-C

$X_i$  : Jarak data obyek ke-i

$C_k$  : Pusat cluster ke-i

*Algoritma K-Medoids* atau dikenal dengan metode PAM (*Partitioning Around Medoids*) merupakan metode yang mewakili cluster dengan data yang sering muncul atau medoid, data yang digunakan adalah data kepolisian daerah di Indonesia yang sudah melalui tahap PCA. Sedangkan *Algoritma Artificial Bee*

*Colony* (ABC) pertama kali dikembangkan oleh Derviş Karaboga pada tahun 2005. ABC melibatkan tiga jenis lebah, yaitu lebah pekerja (*employed bee*), lebah pengamat (*onlooker bee*), dan lebah pengintai (*scout bee*) [15]. Proses pencarian dalam ABC terdiri atas tiga fase utama di setiap siklus iterasi. Pertama, nilai fitness solusi dievaluasi setelah lebah pekerja menghasilkan kandidat solusi awal. Selanjutnya, lebah pengamat memilih wilayah solusi berdasarkan pertukaran informasi fitness dan mengevaluasi kembali nilai fitness pada solusi terpilih. Jika diperlukan, lebah pengintai kemudian dikirim untuk mengeksplorasi solusi baru. Pemilihan solusi oleh lebah pengamat dilakukan secara probabilistik menggunakan rumus probabilitas  $p_i$ .

$$p_i = \frac{fitness_i}{\sum_{n=1}^{SN} fitness_n} \quad (4)$$

Di mana lebah pekerja menilai nilai kesesuaian solusi ke- $i$ . Jumlahnya sebanding dengan nilai  $fitness_i$  keseluruhan solusi di posisi ke- $i$ . Jumlah solusi yang sama dengan jumlah lebah pekerja dilambangkan dengan SN. SN merupakan jumlah solusi yang jumlahnya sama dengan jumlah lebah pekerja. Nilai  $fitness_i$  diperoleh melalui formula dibawah ini:

$$fitness_i = \begin{cases} \frac{1}{1+f_i}, & f_i \geq 0 \\ 1 + abs(f_i), & f_i < 0 \end{cases} \quad (5)$$

dengan  $f_i$  adalah solusi atau fungsi objektif dari  $x_i$  dan  $p_i$  akan berada diantara rentang  $[0,1]$ . Untuk menghasilkan lokasi dari calon solusi baru digunakan rumus berikut:

$$v_{ij} = x_{ij} + \phi_{ij}(x_{ij} - x_{kj}) \quad (6)$$

dengan  $k \in (1,2,...,SN)$  dan  $j \in (1,2,...,D)$  adalah indeks yang dipilih secara acak,  $k$  harus berbeda dengan  $i$ ,  $D$  adalah jumlah parameter optimasi dan  $\phi_{ij}$  angka random berdistribusi uniform dalam rentang  $[-1,1]$ .  $x_{ij}$  adalah sumber solusi baru, dan  $x_{kj}$  adalah sumber solusi tetangga.

$$x_{ij} = x_{minj} + rand(0,1)(x_{maxj} - x_{minj}) \quad (7)$$

dengan  $i = 1, 2, ..., SN$ ,  $SN$  adalah jumlah solusi (jumlah solusi), dan jumlah solusi sama dengan jumlah lebah pekerja.  $x_{ij}$  merupakan solusi ke  $i$ . *Greedy Selection* dilakukan untuk membandingkan  $v_{ij}$  dan  $x_{ij}$  untuk menentukan solusi terbaik, *greedy selection* selalu memilih solusi berikutnya yang menawarkan solusi yang cepat dan jelas.

### 3. HASIL PENELITIAN DAN PEMBAHASAN

#### 3.1. Data Penelitian

**Tabel 2.** Data Penelitian

Kepolisian Daerah	X1	X2	X3	X4	X5	X6	X7	X8
Aceh	664	1544	968	1351	674	1208	643	226
Sumatera Utara	5066	5907	4030	2282	2434	2823	2831	2791
Sumatera Barat	2521	592	2910	1604	967	1741	910	329
Riau	1078	1173	676	541	323	405	624	153
Jambi	1683	1766	612	319	942	467	654	74
Sumatera Selatan	2771	1768	1050	508	1127	282	1237	1239
Bengkulu	810	192	652	466	365	164	83	21
Lampung	1504	1744	271	455	710	1040	707	152
Kep. Bangka Belitung	479	224	126	128	108	3	80	50
Kep. Riau	271	400	416	715	404	345	288	154
Metro Jaya	2180	7214	1796	1324	4236	251	1226	990

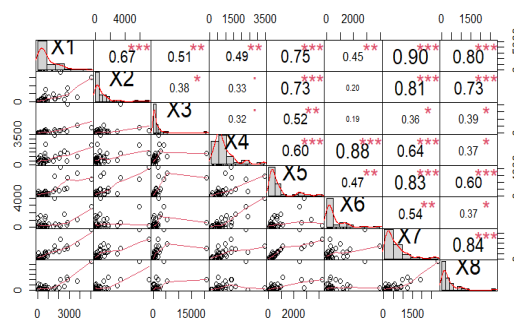
Jawa Barat	3727	1090	5414	960	3345	1049	1429	862
Jawa Tengah	2035	1243	1665	878	365	699	1262	870
DI Yogyakarta	604	290	573	521	510	346	769	187
Jawa Timur	2822	3405	20772	1366	602	693	981	1026
Banten	580	389	810	153	208	143	193	106
Bali	405	665	328	447	418	235	230	6
Nusa Tenggara Barat	309	227	281	649	385	99	56	34
Nusa Tenggara Timur	276	2	493	1069	1533	274	326	484
Kalimantan Barat	684	490	560	970	223	0	406	79
Kalimantan Tengah	288	474	1323	313	113	113	246	7
Kalimantan Selatan	594	1211	494	430	123	132	103	61
Kalimantan Timur	767	2928	1059	584	357	465	174	354
Sulawesi Utara	70	111	250	809	691	0	576	85
Sulawesi Tengah	778	259	2357	2378	667	445	142	44
Sulawesi Selatan	1633	717	2192	332	221	271	126	132
Sulawesi Tenggara	167	65	440	149	744	79	90	9
Gorontalo	15	38	134	711	302	1250	239	34
Sulawesi Barat	16	30	359	438	0	39	6	2
Maluku	99	73	145	213	419	18	95	576
Maluku Utara	8	41	71	74	1	12	107	7
Papua Barat	224	41	74	568	154	163	28	33
Papua	628	41	1920	681	273	413	216	373

Selanjutnya data ini dilakukan standarisasi karena terdapat perbedaan dalam ukuran satuan dari variabel yang digunakan dalam analisis data yaitu adanya perbedaan rentang data yang cukup besar. Sehingga standarisasi digunakan untuk mengubah data menjadi rentang yang lebih kecil. Metode yang sering digunakan dalam menstandarisasi data adalah Z-Score [16].

### 3.2. Uji Korelasi dan Uji Asumsi *Multikolinearitas*

#### 1. Uji Korelasi

Korelasi antar variabel tingkat kejahatan menggunakan *software* R dapat dilihat pada gambar.



Gambar 1. Korelasi

Berdasarkan visualisasi matriks korelasi yang menggambarkan interaksi antara delapan variabel ( $X_1$  hingga  $X_8$ ), bagian segitiga atas menampilkan koefisien korelasi *Pearson*, sementara segitiga bawah berisi *scatter plot*. Nilai korelasi ini mencerminkan intensitas hubungan linear antar pasangan variabel, di mana nilai mendekati 1 atau  $-1$  mengindikasikan korelasi linear kuat, sedangkan nilai mendekati 0 menandakan hubungan lemah atau tidak ada.

Sebagai contoh, variabel  $X_1$  dan  $X_7$  menunjukkan korelasi sebesar 0,90 dengan tingkat signifikansi sangat tinggi (ditandai dengan tiga bintang: \*\*\*), yang menegaskan adanya hubungan positif kuat dan signifikan secara statistik.

Demikian pula, pasangan  $X_5$  dan  $X_6$  memiliki korelasi 0,88\*\*\*, mengindikasikan hubungan linear yang sangat erat. Sebaliknya, korelasi antara  $X_2$  dan  $X_3$  sebesar 0,38 tanpa tanda signifikansi menunjukkan hubungan yang lemah dan tidak signifikan secara statistik.

Selain data numerik, *scatter plot* pada segitiga bawah memberikan wawasan visual tentang pola hubungan. Pola titik yang membentuk garis lurus memperkuat indikasi korelasi, sedangkan penyebaran acak menandakan tidak adanya hubungan yang jelas.

## 2. Uji Asumsi Multikolinearitas

**Tabel 3.** Kriteria Keputusan

	X1	X2	X3	X4	X5	X6	X7	X8
X1	Inf	1.79	1.36	1.31	2.25	1.25	5.50	2.74
X2	1.79	Inf	1.17	1.13	2.11	1.04	2.86	2.12
X3	1.36	1.17	Inf	1.11	1.37	1.04	1.15	1.18
X4	1.31	1.13	1.11	Inf	1.11	1.37	1.04	1.15
X5	2.25	2.11	1.37	1.56	Inf	1.28	3.20	1.57
X6	1.25	1.04	1.04	4.31	1.28	Inf	1.41	1.16
X7	5.50	2.86	1.15	1.68	3.20	1.41	Inf	3.39
X8	2.74	2.12	1.18	1.16	1.57	1.16	3.39	Inf

Berdasarkan Tabel 3.2, nilai VIF yang paling tinggi terdapat pada variabel  $X_7$  sebesar 5.50 dan 3.39 ketika dipasangkan dengan  $X_1$  dan  $X_8$ , sehingga tidak menunjukkan gejala *multikolinearitas* yang serius.

## 3.3. Principal Component Analysis (PCA)

Dalam pengujian asumsi multikolinearitas, terdapat nilai korelasi antarvariabel yang melebihi 0,5, sehingga diperlukan penyederhanaan dimensi melalui transformasi variabel independen menjadi komponen baru yang ortogonal, guna menghilangkan korelasi antarvariabel. Penerapan Analisis Komponen Utama (PCA) mensyaratkan dua asumsi utama yaitu, kecukupan sampel yang divalidasi melalui uji *Kaiser-Meyer-Olkin* (KMO), serta bukti adanya korelasi signifikan antarvariabel yang diuji dengan *Bartlett's Test of Sphericity*.

### a. Asumsi PCA

#### 1. Uji *Kaiser Mayer Olkin* (KMO)

Hasil uji KMO menunjukkan nilai sebesar 0,5732111. Berdasarkan kriteria pada Tabel 2.1, nilai tersebut mengindikasikan bahwa data telah memenuhi syarat kecukupan sampel dan layak digunakan untuk Analisis Komponen Utama (PCA).

#### 2. Uji *Bartlett Sphericity*

**Tabel 3.3.** Nilai Barlett Sphericity

<i>Khi-Squared</i>	<i>Degree Of Freedom (df)</i>	Sig
270.87	28	$2.2 \times 10^{-16}$

Berdasarkan tabel 3.3, terlihat bahwa p-value lebih kecil dari  $\alpha$  (0,05) sehingga memberikan keputusan tolak  $H_0$ . Artinya terdapat korelasi yang signifikan antar variabel sehingga perlu untuk dilakukan analisis PCA.

b. Hasil Analisis Komponen Utama

Hasil Analisis Komponen Utama (PCA) dengan perangkat lunak R menghasilkan nilai-nilai seperti yang tercantum pada tabel berikut. Tabel 3.4 menyajikan nilai eigenvalue untuk setiap komponen yang terbentuk, di mana *eigenvalue* mencerminkan proporsi varians yang dijelaskan oleh masing-masing komponen utama serta kontribusinya dalam menangkap keragaman data asli. Penentuan jumlah komponen yang dipertahankan mengacu pada kriteria *eigenvalue* > 1. Dari Tabel 3.4, terlihat bahwa hanya komponen 1 dan komponen 2 yang memiliki *eigenvalue* di atas 1, yang berarti kedua komponen tersebut secara bersama-sama telah cukup mewakili varians dari delapan variabel asal. Dengan demikian, dua komponen ini dianggap representatif untuk menggambarkan delapan jenis kejahatan pada setiap kepolisian daerah di Indonesia.

**Tabel 4.** Nilai *Eigen Value*

Komponen	Nilai Eigen Value
1	5.02869789
2	1.28917450
3	0.77866580
4	0.41660101
5	0.25292548
6	0.12789463
7	0.09443769
8	0.01160301

c. Persamaan Komponen Utama

Langkah selanjutnya adalah menentukan *eigenvector* atau koefisien dari komponen utama, yang akan digunakan untuk menyusun persamaan komponen utama. Hasil perhitungan koefisien komponen utama dengan menggunakan perangkat lunak R disajikan pada Tabel 3.5 berikut.

**Tabel 5.** Koefisien Komponen Utama

Variabel	PC1	PC2
X1	0.40	-0.15
X2	0.35	-0.36
X3	0.25	-0.17
X4	0.32	0.57
X5	0.39	-0.04



X6	0.28	0.65
X7	0.43	-0.06
X8	0.37	-0.25

Dari tabel 3.5 diatas, nilai dari PC1 dan PC2 untuk masing-masing variabel dapat membentuk dua persamaan sebagai berikut:

$$PC1 = 0,40X1 - 0,35X2 - 0,25X3 - 0,32X4 - 0,39X5 - 0,28X6 - 0,43X7 - 0,37X8$$

$$PC2 = 0,40X1 - 0,36X2 - 0,17X3 - 0,57X4 - 0,04X5 - 0,65X6 - 0,06X7 - 0,25X8$$

Persamaan tersebut selanjutnya digunakan untuk menghitung nilai atau score dari kedua Komponen Utama (PC1 dan PC2). Berikut adalah hasil perhitungan score PC1 dan PC2 untuk setiap provinsi dengan menggunakan perangkat lunak R.

**Tabel 6.** *Principal Component Score*

Kepolisian Daerah	PC1	PC2
Aceh	0.176684455	0.866608391
Sumatera Utara	7.736492578	-0.308743397
Sumatera Barat	1.389199326	1.271225675
Riau	-0.574122909	-0.300911339
Jambi	0.287020097	0.012632107
Sumatera Selatan	1.563139495	-1.364808987
Bengkulu	-1.203827635	-0.339651561
Lampung	-0.279951094	-0.750065987
Kep. Bangka Belitung	-1.892208503	-0.49535697
Kep. Riau	-1.20338441	0.121401277
Metro Jaya	4.428778056	-1.984186482
Jawa Barat	3.188744417	-0.568137047
Jawa Tengah	0.193865738	-0.31538043
DI Yogyakarta	-1.250159624	-0.185277985
Jawa Timur	4.012486282	-1.576494453
Banten	-1.461846539	-0.453144838
Bali	-1.417924648	-0.026932254
Nusa Tenggara Barat	-1.193259926	-0.106966326
Nusa Tenggara Timur	-0.505946443	1.226322805
Kalimantan Barat	-1.03927094	0.016152458
Kalimantan Tengah	-1.811202848	-0.431610074
Kalimantan Selatan	-1.149431558	-0.556816282
Kalimantan Timur	-0.20597896	-0.865449177
Sulawesi Utara	-0.804164608	-0.205617477
Sulawesi Tengah	0.456418141	2.190405839
Sulawesi Selatan	3.465345896	4.446722863

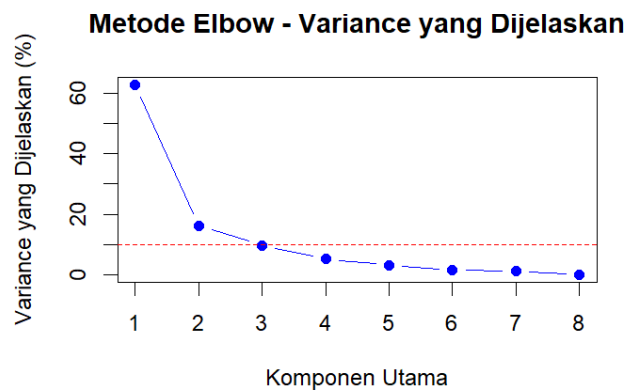


Sulawesi Tenggara	-1.665266049	0.392858148
Gorontalo	-1.254322299	0.983937771
Sulawesi Barat	-1.755167022	0.10131435
Maluku	-1.47193611	-0.350299721
Maluku Utara	-2.142301146	-0.427769748
Papua Barat	-1.615163377	-0.038093771
Papua	-1.001337834	0.022132622

Berdasarkan tabel, diperoleh nilai *principal component score* (PC) untuk seluruh kepolisian daerah di Indonesia. Nilai-nilai inilah yang kemudian digunakan sebagai dasar pengelompokan kepolisian daerah di Indonesia.

d. Penentuan Jumlah *K Cluster* Optimal

Dalam penelitian ini, jumlah kluster optimal (*K*) ditentukan menggunakan metode *Elbow*. Kluster yang paling tepat dapat diidentifikasi pada titik siku yang terbentuk pada grafik berikut.



**Gambar 2.** *Elbow*

Dari grafik tersebut, titik siku terlihat berada di antara kluster 2 dan kluster 3. Titik siku ini menandakan penurunan nilai *Sum of Squared Error* (SSE) yang signifikan antara kedua kluster tersebut, diikuti oleh penurunan yang jauh lebih kecil pada kluster-kluster berikutnya. Oleh karena itu, jumlah kluster optimal untuk pengelompokan menggunakan metode *K-Medoids* berbasis *Artificial Bee Colony* (ABC) adalah 3 kluster.

### 3.4. Pembentukan *Cluster* Menggunakan *K-Medoids* dengan Algoritma *Artificial Bee Colony*

Teknik pengelompokan partisi *K-Medoids Artificial Bee Colony* bertujuan mengidentifikasi *k* kluster yang paling representatif terhadap objek-objek dalam dataset.

1.  $K=3$
2. Kemudian kita menentukan pusat *cluster* (*centroid*) secara acak.

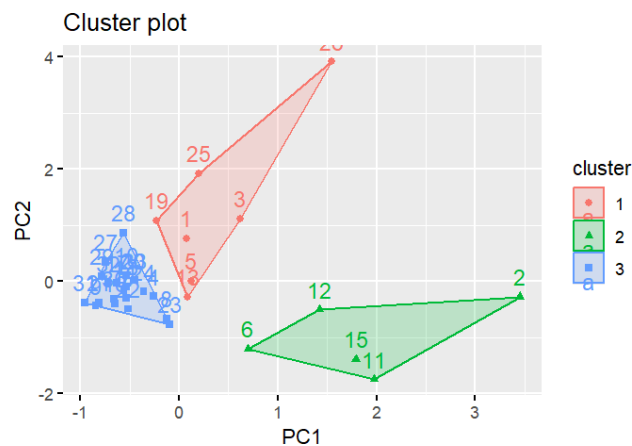
**Tabel 7.** *Centroid Awal K-Medoids Artificial Bee Colony*

Variabel	C1	C2	C3
$PC_1$	0.1766845	4.0124863	-1.2501596
$PC_2$	0.8666084	-1.5764945	-0.1852780

3. Selanjutnya, dihitung jarak antara setiap objek dan pusat klaster (*centroid*) masing-masing menggunakan jarak *Euclidean*. Berikut adalah contoh perhitungan manual jarak antara Kepolisian Daerah Aceh (Objek 1) dengan centroid awal C1, C2, dan C3 (mewakili klaster 1, 2, dan 3) sebagaimana tercantum pada Tabel 3.7.
4. Dari hasil perhitungan, diperoleh jarak Kepolisian Daerah Aceh terhadap ketiga centroid. Nilai terkecil terdapat pada jarak dengan C1, yang menunjukkan bahwa Kepolisian Daerah Aceh paling dekat dengan centroid klaster 1. Dengan demikian, wilayah ini memiliki karakteristik yang paling mirip dengan klaster 1 dan dialokasikan sebagai anggota klaster tersebut.

Proses serupa diterapkan pada seluruh kepolisian daerah, mulai dari Sumatera Utara hingga Papua, dengan membandingkan jarak terhadap *centroid* masing-masing klaster. Setiap objek dialokasikan ke klaster terdekat (dengan jarak terkecil), diikuti perhitungan total jarak keseluruhan. Tahap ini merupakan iterasi pertama dalam algoritma *K-Medoids Artificial Bee Colony*.

Berdasarkan data yang ada saat ini, anggota cluster untuk setiap nama kepala keluarga yang tergabung dalam satu cluster merupakan wilayah yang ciri-cirinya seragam. Adapun hasil *K-Medoids Artificial Bee Colony* yaitu Cluster 1 terdiri dari 7 kepolisian daerah, Cluster 2 terdiri dari 5 kepolisian daerah, dan Cluster 3 terdiri dari 21 kepolisian daerah (per kelompok).



**Gambar 3.** Hasil Pengelompokkan *K-Medoids Artificial Bee Colony*

### 3.5. Hasil Cluster menggunakan *K-Medoids Artificial Bee Colony*

**Tabel 8.** Hasil *K-Medoids Artificial Bee Colony*

Cluster	Jumlah	Anggota
1	7	Aceh, Sumatera Barat, Jambi, Jawa Tengah, Nusa Tenggara Timur, Sulawesi Tengah dan Sulawesi Selatan.
2	5	Sumatera Utara, Sumatera Selatan, Metro Jaya, Jawa Barat dan Jawa Timur.
3	21	Riau, Bengkulu, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, Yogyakarta, Banten, Bali, Nusa Tenggara Barat, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Sulawesi Utara, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat dan Papua.

### 3.6. Validasi Clustering

Validasi hasil pengelompokan *K-Medoids Artificial Bee Colony* dilakukan dengan menggunakan *Silhouette Index* (SI), yang mengukur kualitas kluster dengan rentang nilai antara -1 hingga +1. Nilai SI di atas 0,5 menunjukkan pengelompokan yang baik, di mana objek dalam satu kluster sangat kompak dan terpisah dengan jelas dari kluster lain. Nilai antara 0,2 hingga 0,5 mencerminkan pengelompokan yang cukup baik meskipun terdapat sedikit tumpang tindih antar-kluster, sedangkan nilai di bawah 0,2 mengindikasikan hasil yang buruk karena banyak objek lebih cocok berada di kluster lain. Hasil evaluasi *Silhouette Index* untuk metode *K-Medoids Artificial Bee Colony* disajikan dalam tabel berikut.

**Tabel 9.** Nilai Indeks Validitas *Cluster Silhouette Index* (SI) Metode *K-Medoids Artificial Bee Colony*

Cluster	Metode	Nilai SI
3	<i>K-Medoids Artificial Bee Colony</i>	0.3870265

Tabel di atas menyajikan hasil evaluasi *Silhouette Index* (SI) pada metode *K-Medoids* berbasis *Artificial Bee Colony* (ABC) untuk pengelompokan. Dengan nilai SI sebesar 0,387, hasil ini mengindikasikan bahwa pengelompokan termasuk dalam kategori cukup baik.

### 3.7. Interpretasi

**Tabel 10.** Nilai Rata-Rata *Principal Component* Tiap Cluster

Variabel	Cluster 1	Cluster 2	Cluster 3
Kejahatan Terhadap Nyawa	0,782749368	1,864695888	-0,198482118
Kejahatan Terhadap Fisik	0,615602762	1,708728437	-0,139992106
Kejahatan Terhadap Kesusilaan	0,34274824	1,366153614	-0,039276924
Kejahatan Terhadap Kemerdekaan Orang	0,549180087	0,59944732	-0,17677411
Kejahatan Terhadap Hak Milik dengan Penggunaan Kekerasan	0,623266804	1,848030163	-0,072607817
Kejahatan Terhadap Hak Milik dengan Tanpa Penggunaan Kekerasan	0,220511217	0,42408192	-0,210935668
Kejahatan Terkait Narkotika	0,782448833	1,803793473	-0,20573769
Kejahatan Terkait Penipuan, Penggelapan, Korupsi	0,340541265	1,866570117	-0,196053875

Tabel tersebut menyajikan nilai rata-rata *Principal Component* setiap variabel pada masing-masing kluster yang terbentuk, mencakup berbagai jenis kejahatan mulai dari Kejahatan Terhadap Nyawa, Kejahatan Terhadap Fisik, hingga Penipuan, Penggelapan, dan Korupsi. Pada Kluster 1, rata-rata Kejahatan Terhadap Nyawa tercatat sebesar 0,782749368, lebih rendah dibandingkan Kluster 2 (1,864695888) dan Kluster 3 (-0,189482118), menandakan bahwa kluster pertama memiliki tingkat kejahatan jenis ini yang relatif rendah. Pola serupa terlihat pada variabel Kejahatan Terhadap Fisik, dengan nilai tertinggi di Kluster 2 (1,707287348) dan terendah di Kluster 3 (-0,13992106).

Secara keseluruhan, perbandingan rata-rata variabel ini menggambarkan perbedaan karakteristik yang jelas antar-kluster. Kluster 3 cenderung memiliki tingkat kejahatan yang lebih rendah secara umum, Kluster 2 menonjol dengan nilai tinggi pada sebagian besar jenis kejahatan terutama terhadap nyawa dan kejahatan terhadap penipuan, penggelapan, korupsi. Sedangkan kluster 1 menunjukkan tingkat kejahatan sedang dan relatif merata. Kluster 3 menunjukkan nilai rendah di seluruh variabel, nilai yang paling rendah ditunjukkan oleh kejahatan terhadap hak milik dengan penggunaan kekerasan dan kejahatan terkait narkotika. Kluster ini mencerminkan wilayah dengan kondisi relatif aman dilihat dari tingkat kejahatan yang sangat rendah dibandingkan dengan kluster lain. Tabel ini memberikan wawasan mendalam tentang profil masing-masing kluster, yang dapat menjadi dasar analisis lanjutan terkait pola dan faktor pendorong distribusi kejahatan di berbagai wilayah.

Pada bagian ini dijelaskan hasil-hasil penelitian dan dijelaskan secara komprehensif. Hasil penelitian dapat disajikan dalam bentuk gambar, grafik, tabel dan lain-lain yang memudahkan pembaca memahami

hasil penelitian. Pembahasan dapat dilakukan dalam beberapa sub-bagian (sub-bab). Sangat disarankan untuk memberikan perbandingan dengan penelitian penelitian yang pernah ada.

#### 4. KESIMPULAN

Indonesia. Pengelompokan menggunakan metode *K-Medoids Artificial Bee Colony* menghasilkan 3 klaster optimal, yaitu Klaster 1 dengan 7 kepolisian daerah, Klaster 2 dengan 5 kepolisian daerah, dan Klaster 3 dengan 21 kepolisian daerah. Evaluasi melalui *Silhouette Index* (SI) sebesar 0,387 menunjukkan bahwa hasil pengelompokan termasuk kategori struktur lemah, yang artinya struktur klaster terbentuk namun dengan tingkat pemisahan yang lemah (*Weak Separation*) . Klaster 1 cenderung memiliki nilai rendah pada sebagian besar variabel kejahatan, Klaster 2 menonjol dengan nilai tinggi pada Kejahatan Terhadap Nyawa dan Kejahatan Terhadap Fisik yang mengindikasikan tingkat kejahatan serius yang lebih tinggi, sedangkan Klaster 3 menampilkan nilai rendah secara keseluruhan yang menunjukkan klaster ini memiliki kondisi relatif aman terhadap kejahatan. Temuan ini memberikan gambaran perbedaan profil kejahatan antar-klaster yang dapat dimanfaatkan untuk analisis lanjutan. Kelemahan dari penelitian ini terdapat pada hasil SI yang mengindikasikan bahwa klaster yang terbentuk tergolong struktur yang lemah oleh karena itu untuk penelitian selanjutnya bisa lebih baik jika menggunakan variabel yang lebih banyak dan kompleks agar mendapatkan hasil yang lebih maksimal, dan juga bisa dicoba menggunakan uji validasi cluster lain seperti Calinski-Harabasz untuk nilai cluster yg maksimal.

#### DAFTAR PUSTAKA

- [1] R. E. Caraka *et al.*, “Micro, small, and medium enterprises’ business vulnerability cluster in Indonesia: An analysis using optimized fuzzy geodemographic clustering,” *Sustainability*, vol. 13, no. 14, p. 7807, 2021.
- [2] B. Probierz, J. Kozak, and A. Hrabia, “Clustering of scientific articles using natural language processing,” *Procedia Comput. Sci.*, vol. 207, pp. 3449–3458, 2022.
- [3] A. T. R. Dani, S. Wahyuningsih, and N. A. Rizki, “Pengelompokan Data Runtun Waktu menggunakan Analisis Cluster: Studi Kasus: Nilai Ekspor Komoditi Migas dan Nonmigas Provinsi Kalimantan Timur Periode Januari 2000-Desember 2016,” *Ekspansional*, vol. 11, no. 1, pp. 29–38, 2020.
- [4] A. T. R. Dani, S. Wahyuningsih, and N. A. Rizki, “Penerapan Hierarchical Clustering Metode Agglomerative pada Data Runtun Waktu,” *Jambura J. Math.*, vol. 1, no. 2, pp. 64–78, 2019.
- [5] I. Ayuningtias and N. Debataraja, Naomi Nessyana Imro’ah, “Analisis cluster non-hirarki dengan metode k-modes,” *Bimaster Bul. Ilm. Mat. Stat. dan Ter.*, vol. 8, no. 4, 2019.
- [6] M. A. Nahdliyah, T. Widiharih, and A. Prahutama, “Metode K-Medoids Clustering dengan Validasi Silhouette Index dan C-Index (Studi Kasus Jumlah Kriminalitas Kabupaten/Kota di Jawa Tengah Tahun 2018),” *J. Gaussian*, vol. 8, no. 2, pp. 161–170, 2019.
- [7] P. K. Panigrahi, S. Ghosh, and D. R. Parhi, “Comparison of GSA, SA and PSO based intelligent controllers for path planning of mobile robot in unknown environment,” *Int. J. Electr. Comput. Eng.*, vol. 8, no. 10, pp. 1633–1642, 2014.
- [8] C. A. Asrat, Makkulau, and I. Yahya, “Perbandingan Metode Principal Componen Analysis (PCA) dan Partial Least Square (PLS) dalam Penanganan Multikolinearitas pada Kasus Kemiskinan di Provinsi Sulawesi Tenggara Tahun 2023,” *Arus J. Sains dan Teknol.*, vol. 3, no. 1, pp. 68–82, 2025.
- [9] A. Islamiyati, S. Sahriman, and S. Oktoni, “Studi Longitudinal Pada Analisis Data Gula Darah Pasien Diabetes melalui Principal Component Analysis,” *Jambura J. Math.*, vol. 4, no. 1, pp. 41–49, 2022.
- [10] A. N. Nitasari, F. N. Salsabila, D. T. Ramadhanty, M. R. Anggriawan, D. Amelia, and E. Mardianto, M Fariz Fadillah Ana, “Reduksi Faktor-Faktor yang Mempengaruhi Kualitas Air Hujan dengan Metode Analisis Komponen Utama,” *Zeta-Math J.*, vol. 8, no. 1, pp. 7–15, 2023.
- [11] M. N. Van Delsen and S. Wattimena, Abraham Saputri, “Penggunaan metode analisis komponen utama untuk mereduksi faktor-faktor inflasi di Kota Ambon,” *Barekeng J. Ilmu Mat. dan Terap.*,

- vol. 11, no. 2, pp. 109–118, 2017.
- [12] B. T. El Haqq, A. Antika, and S. P. Wulandari, “Analisis Faktor-faktor Volume Ekspor Hasil Perikanan Menurut Provinsi di Indonesia Tahun 2021 menggunakan Analisis Faktore,” *Zool. J. Ilmu Peternakan, Ilmu Perikanan, Ilmu Kedokt. Hewan*, vol. 3, no. 1, pp. 01–18, 2025.
- [13] V. B. Samosir, A. M. Widodo, N. Anwar, B. A. Sekti, and N. Erzed, “Identifikasi outlier menggunakan teknik data mining clustering untuk analisis data tracer study pada fakultas ilmu komputer universitas esa unggul,” *IKRA-ITH Inform. J. Komput. dan Inform.*, vol. 8, no. 1, pp. 162–174, 2024.
- [14] D. Ayu, I. C. Dewi, and K. Pramita, “Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali,” *J. Matrix*, vol. 9, no. 3, p. 102, 2019.
- [15] M. Furqan, Y. R. Nasution, and K. Khairunnisa, “Application of Artificial Bee Colony Algorithm to Optimize the Shortest Route to Distribute Clean Water Pipes,” *JOMLAI J. Mach. Learn. Artif. Intelligence*, vol. 1, no. 2, pp. 125–132, 2022.
- [16] I. M. Karo Karo and H. Hendriyana, “Klasifikasi Penderita Diabetes menggunakan Algoritma Machine Learning dan Z-Score,” *J. Teknol. Terpadu*, vol. 8, no. 2, pp. 94–99, Dec. 2022, doi: 10.54914/jtt.v8i2.564.