

# Analisis *Cluster* Kriminalitas 2019 Menurut Kepolisian Daerah dengan *K-Means* dan *Hierarchical*

Muhammad Jamie Rofie Quality (2006473812)

Departemen Ilmu Ekonomi, Fakultas Ekonomi dan Bisnis, Universitas Indonesia, Jl. Prof. Dr. Sumitro  
Djojohadikusumo, Kota Depok, Jawa Barat

## 1. Latar Belakang

Kriminalitas adalah salah satu problem sosial yang berpengaruh negatif pada kesejahteraan masyarakat, baik secara material maupun batin. Selain penanganan represif yang secara langsung menangani kejahatan di lapangan, investigasi kriminalitas juga diperlukan sebagai strategi preventif untuk meredamnya. Banyak faktor yang digunakan untuk menganalisis perilaku kriminalitas, mulai dari ekonomi, politik, demografi, hingga spasial. *Institute of Economics and Peace* (IEP) memberi Indonesia peringkat ke-47 dalam indeks kedamaian pada tahun 2022,<sup>[1]</sup> turun dari tahun sebelumnya yang berada pada posisi ke-42. Dilansir dari Numbeo, Indonesia menempati peringkat ke-52 dari 136 negara dalam Indeks Kriminalitas dengan skor 46,01 pada 2019 dan turun menjadi peringkat ke-73 pada pertengahan 2022.<sup>[2]</sup>

Mengetahui pentingnya prevensi kriminalitas, maka penting untuk melakukan investigasi kriminalitas dalam skala makro. Ada beberapa jenis kriminalitas sesuai dengan bagaimana masing-masing negara melakukan kategorisasi. Badan Pusat Statistik membagi kriminalitas menjadi 9 klasifikasi dan 22 jenis. Sembilan klasifikasi tersebut terbagi berdasarkan target, tingkat keseriusan, dan cara dilakukannya kejahatan yang berkaitan: nyawa, fisik, kesusilaan, kemerdekaan individu, hak milik (terbagi menjadi dua, dengan dan tanpa kekerasan), narkoba, penipuan dan korupsi, dan ketertiban umum.<sup>[3]</sup> Dalam makalah ini, penulis ingin menganalisis kriminalitas ini dari klasifikasi kejahatan itu sendiri dan membentuk klaster ke seluruh provinsi (Kepolisian Daerah) di Indonesia.

## 2. Data

Analisis ini menggunakan data yang dipublikasikan oleh Badan Pusat Statistik dengan judul “STATISTIK KRIMINAL 2021”.<sup>[4]</sup> Variabel yang digunakan adalah masing-masing klasifikasi kejahatan yang berdasarkan Kitab Undang-Undang Hukum Pidana (KUHP) RI dan *The International Classification of Crime for Statistical Purposes* (ICCS). Data yang

digunakan bersifat *cross-section* dengan pembagian berdasarkan provinsi di Indonesia atau kepolisian daerah (Polda). Karena data kejahatan bersifat jumlah kejadian yang ada, maka peneliti melakukan pembobotan berdasarkan populasi masing-masing provinsi dan menghasilkan tingkat kejahatan yang terjadi pada setiap 1.000 penduduk provinsi. Berikut adalah variabel yang digunakan dalam analisis klaster.

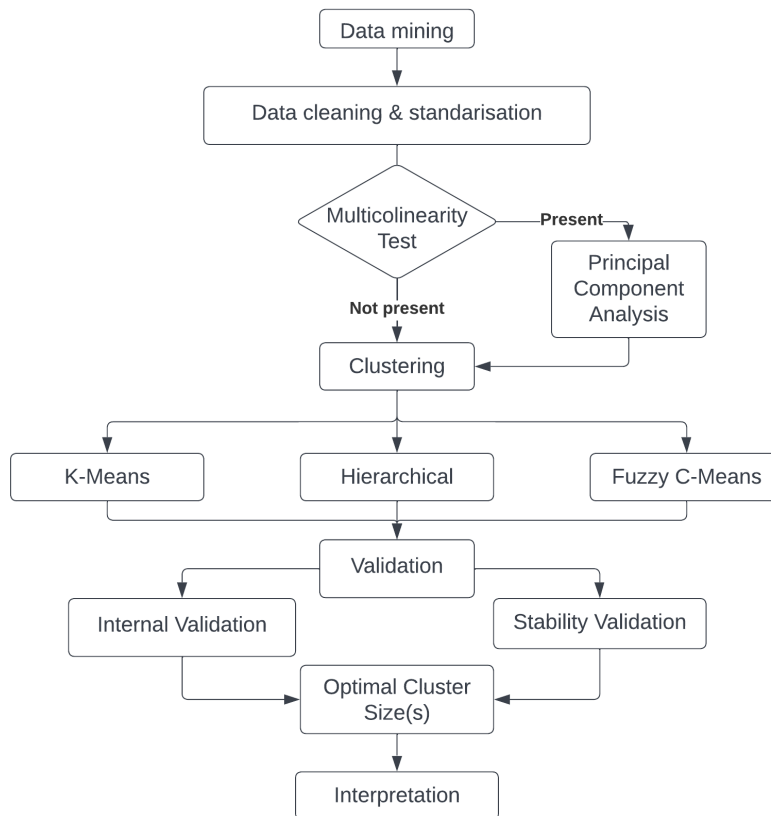
Tabel 1. Daftar Variabel dan Deskripsi

Variabel	Deskripsi
polda	Provinsi atau satuan Kepolisian Daerah terjadinya kriminalitas. Merupakan variabel <i>string</i> .
murder	Kejahatan terhadap nyawa
assault	Kejahatan terhadap fisik/badan
rape	Kejahatan terhadap kesusilaan. Meliputi pemerkosaan dan pencabulan
abduction	Kejahatan terhadap kemerdekaan orang
theftv	Kejahatan terhadap hak milik dengan penggunaan kekerasan
thiftwv	Kejahatan terhadap hak milik tanpa penggunaan kekerasan
drugs	Kejahatan terkait narkoba
fnc	Kejahatan terkait penipuan, penggelapan, dan korupsi ( <i>fraud and corruption</i> )
capo	Kejahatan terhadap ketertiban umum ( <i>Crime against public order</i> )

## 3. Metode

Analisis *cluster* ini menggunakan tiga metode *clustering* serta membandingkan hasil ketiganya: *K-Means*, *Hierarchical*, dan Fuzzy C-Means. Validasi internal juga akan dilakukan untuk memberikan penilaian akan masing-masing metode yang ada. Alur penelitian digambarkan pada diagram pada **gambar 1**.

KMO adalah salah satu metode pengujian



Gambar 1. Alur Penelitian

### 3.1. Clustering

Analisis *clustering* adalah teknik yang umum digunakan dalam analisis data dengan mengelompokkan objek yang mirip pada kelompok yang berbeda, atau membagi data ke subset-subset, sehingga data pada setiap subset merujuk pada ukuran jarak yang telah ditentukan.<sup>[5]</sup> *Clustering* merupakan salah satu jenis analisis yang dapat dilakukan pada studi *machine learning*. Selanjutnya, istilah klaster dan kelompok akan digunakan secara bergantian.

### 3.2. Principal Component Analysis (PCA) <sup>[6]</sup>

PCA adalah salah satu cara mereduksi *dimensionality* dari dataset tanpa mengurangi variabilitasnya. Hal ini dapat dilakukan dengan menggabungkan beberapa variabel secara linier berdasarkan variasi totalnya. Beberapa gabungan ini nantinya akan menciptakan komponen baru. Komponen ini nantinya akan menggantikan fungsi variabel dalam data awal.

### 3.3. Kaiser-Meyer-Olkin Statistics (KMO) <sup>[7]</sup>

kecukupan sampel (*sample adequacy*) dalam statistika. Uji ini mengukur korelasi antar variabel dan juga menguji apakah analisis faktor perlu dilakukan. Nilai KMO yang signifikan lebih kecil dari 0,50 (atau 0,60) menandakan tolak hipotesis bahwa analisis faktor menggunakan data terkait dapat dilakukan. Apabila nilainya lebih tinggi dari 0,50 atau 0,60, maka gagal tolak hipotesis bahwa analisis faktor dapat dilakukan.

### 3.4. Bartlett's Test of Sphericity <sup>[8]</sup>

Tes Bartlett digunakan untuk menguji apakah matriks korelasi memiliki kesamaan dengan matriks identitas atau menguji apakah setiap variabel berkorelasi dengan variabel itu sendiri. Hipotesis dari tes Bartlett adalah sebagai berikut:

$H_0$ : Matriks Korelasi = Matriks identitas

$H_A$ : Matriks Korelasi  $\neq$  Matriks identitas

### 3.5. K-Means

K-Means adalah salah satu metode *clustering*. Metode ini menggunakan algoritma yang menetapkan setiap poin ke klaster dengan *centroid* terdekat.<sup>[9]</sup> K-

Means merupakan salah satu metode *partitional clustering*. Metode ini akan menetapkan titik secara acak sesuai dengan jumlah kluster yang telah ditentukan. Titik ini disebut sebagai *centroid* atau titik pusat. Masing-masing poin atau data akan disesuaikan dengan *centroid* terdekat sesuai dengan jarak masing-masing data dengan *centroid* dan membentuk kluster. Algoritma K-Means akan menyesuaikan *centroid* yang baru, berdasarkan anggota kluster yang telah dibentuk. *Centroid* baru juga dapat mengubah anggota kluster, tergantung dari jaraknya ke data. Proses pengubahan *centroid* dan penyesuaian anggota kluster akan terus diulang hingga mencapai *convergence*.

### 3.6. Hierarchical Clustering <sup>[10]</sup>

Metode kluster *hierarchical* mengelompokkan data dengan menggabungkan kluster secara progresif sesuai dengan jarak masing-masing data. Ada beberapa metode dalam penentuan jarak masing-masing data (*linkage*):

#### 3.6.1. Single Linkage

Jarak antar kluster merupakan selisih minimum pada setiap anggota kluster tertentu. Jarak kluster A dengan B didefinisikan sebagai berikut:

$$\min \{d(x, y): x \in A, y \in B\}$$

keterangan:  $d(x, y)$  adalah jarak data ke- $x$  dengan data ke- $y$ .

#### 3.6.2. Complete Linkage

Jarak antar kluster adalah selisih maksimum pada setiap anggota kluster. Jarak kluster A dengan B dinotasikan sebagai berikut:

$$\max \{d(x, y): x \in A, x \in B\}$$

#### 3.6.3. Average Linkage

Jarak antar kluster adalah rata-rata jarak antar setiap anggota kluster. Jarak kluster A dengan B dinotasikan sebagai berikut:

$$\frac{1}{|A||B|} \sum_{x \in A} \sum_{y \in B} d(x, y)$$

Keterangan:  $|A|$  adalah jumlah data pada kluster A.

#### 3.6.4. Ward's Method

Metode Ward tidak menggunakan jarak kluster sebagai dasar penentuan kelompok, tetapi menggunakan total *sum of squares* (SSE) dalam kluster untuk

menentukan kelompok yang akan digabungkan. SSE ini dinotasikan sebagai berikut:

$$SSE = \sum_{x=1}^K \sum_{y=1}^{n_i} (y_{xy} - \bar{y}_x)^2$$

Keterangan:  $y_{xy}$  adalah data ke- $j$  pada kluster ke- $i$  dan  $n_i$  adalah jumlah objek pada kluster ke- $i$ .

### 3.7. Fuzzy C-Means Clustering (FCM) <sup>[11]</sup>

Proses *clustering* menggunakan metode FCM merupakan pengembangan dari metode *Hard C-Mean* (HCM). Pada metode HCM, setiap data diberi label anggota baik 0 atau 1. Namun, pada FCM, nilai itu direpresentasikan dengan derajat keanggotaan antara 0 hingga 1. Apabila data tertentu memiliki derajat yang tinggi pada kluster  $i$ , hal itu menandakan bahwa data tersebut cocok di kluster  $i$ .

### 3.8. Validasi Internal Kluster

Validasi adalah salah satu langkah dari analisis *clustering* yang menilai seberapa baik kluster-kluster yang telah dibentuk. Validasi internal menggunakan informasi yang ada pada struktur data terkait, tanpa menggunakan informasi eksternal. Dua hal yang menjadi tolok ukur validasi internal adalah *compactness* dan *separation*. <sup>[12]</sup>

#### 3.8.1. Compactness

Validasi internal mengukur dekatnya kemiripan setiap data pada suatu kluster. *Compactness* juga terkadang disebut sebagai *cohesiveness*. Hal ini dapat diukur dengan variansnya: semakin kecil varians menandakan *compactness* yang baik, *vice versa*.

#### 3.8.2. Separation

Validasi internal mengukur jelasnya perbedaan atau terpisahnya satu kluster dari kluster lainnya. Pengukuran ini dapat menggunakan jarak *pairwise* hingga *density*.

Pengukuran kedua kriteria tersebut dapat menggunakan beberapa indeks, diantaranya yang paling sering digunakan ialah indeks Dunn, Silhouette, Connectivity, dan Davies-Bouldin.

#### 3.8.3. Indeks Dunn <sup>[13]</sup>

Indeks Dunn menggunakan jarak pasangan data pada kluster yang berbeda sebagai parameter *separation* dan menggunakan diameter maksimum kluster-kluster yang diciptakan sebagai parameter *compactness*. Secara

matematis Indeks Dunn dapat didefinisikan sebagai berikut:

$$\min_i \left\{ \min_j \left( \frac{\min_{x \in C_i, y \in C_j} d(x, y)}{\max_k \left\{ \max_{x, y \in C_k} d(x, y) \right\}} \right) \right\}$$

Nilai indeks Dunn yang semakin tinggi menandakan jumlah kluster yang optimal.

### 3.8.4. Indeks Silhouette <sup>[14]</sup>

Indeks *Silhouette* mengukur baiknya model *clustering* dengan perbedaan dari jarak antar dan dalam kluster berpasangan. Indeks *Silhouette* dapat dinotasikan secara matematis sebagai berikut:

$$\frac{1}{NC} \sum_i \left\{ \frac{1}{n_i} \sum_{x \in C_i} \frac{b(x) - a(x)}{\max [b(x) - a(x)]} \right\}$$

Nilai indeks *Silhouette* yang semakin tinggi menandakan jumlah kluster yang semakin optimal.

### 3.8.5. Indeks Connectivity <sup>[15]</sup>

Indeks *Connectivity* dibagi lagi menjadi *Intracluster Connectivity* yang mengukur *compactness* dan *Intercluster Connectivity* yang mengukur *separation*. *Intracluster Connectivity* dapat dinotasikan sebagai berikut:

$$\frac{\sum_{a,b}^P CADJ(a, b) : p_a, p_b \in C_k}{\sum_{a,b}^P CADJ(a, b) : p_a \in C_k}$$

*Intracluster Connectivity* dapat dinotasikan sebagai berikut:

$$\left\{ \frac{0, CONN(C_K, C_I)}{\sum_{a,b}^P \{ CONN(a, b) : p_a \in C_k, p_b \in C_K \}} \right\}$$

Nilai indeks *connectivity* yang semakin kecil menandakan jumlah kluster yang optimal.

Keterangan:  $K$  adalah jumlah kluster,  $P$  adalah jumlah prototip,  $Conn(a, b)$  adalah matriks simetris, dan  $CADJ(a, b)$  adalah matriks dekat kumulatif yang tidak simetris.

### 3.8.6. Within dan Between Cluster Sum of Square <sup>[16]</sup>

*Within Cluster Sum of Squares* (WSS) adalah salah satu cara penentuan jumlah kluster dengan validasi *compactness*. Perhitungan WSS adalah sebagai berikut:

$$\sum_i \sum_{x \in C_i} (x - m_i)^2$$

*Between Sum of Squares* (BSS), di lain sisi, memvalidasi *separation* antar kluster. Perhitungan BSS dapat dinotasikan secara matematis sebagai berikut:

$$\sum_i |C_i| (m - m_i)^2$$

Jumlah kluster yang optimal dapat dinilai menggunakan *Elbow method* atau dengan melihat tekukan yang signifikan pada kurva setiap nilai WSS. Selain itu, dapat juga menggunakan *Silhouette method* yang menggunakan koefisien *Silhouette* dan menentukan nilai koefisien tertingginya. Metode ini dapat dinotasikan sebagai berikut:

$$s = 1 - \left( \frac{WSS}{BSS} \right), \text{ jika } WSS > BSS$$

$$s = \left( \frac{BSS}{WSS} \right) - 1, \text{ jika } WSS < BSS$$

## 4. Hasil dan Pembahasan

### 4.1. Analisis Faktor

Analisis faktor diperlukan untuk menentukan apakah sampel yang digunakan cukup dan tidak memiliki multikolinearitas. Uji kecukupan (*adequacy*) dilakukan menggunakan uji Kaiser-Meyer-Olkin (KMO). Hasil uji menggunakan KMO ada pada tabel 2. Pada tabel tersebut didapatkan nilai MSA dari uji KMO sebesar 0,68. Nilai ini menandakan bahwa sampel dapat digunakan untuk analisis faktor, karena berada di atas 0,60.

Tes multikolinearitas pada sampel dilakukan dengan uji Bartlett. Hasil dari uji Bartlett ada pada tabel 2. Dapat dilihat bahwa *p-value* cukup rendah sehingga menandakan terdapat cukup bukti bahwa terdapat multikolinearitas pada sampel yang digunakan. *Principal component analysis* dapat dilakukan untuk mereduksi variabel-variabel dan membuat analisis *clustering* dapat dilakukan.

Hasil dari analisis PCA dapat dilihat pada tabel 3. Proses analisis faktor menggunakan PCA menghasilkan nilai *eigenvalue* yang merupakan standar deviasi masing-masing komponen. Dari ke-9 komponen yang digenerasi, sebanyak 3 komponen memiliki *eigenvalue*, yaitu yang memiliki standar deviasi lebih dari 1,00. Hal ini menandakan bahwa ada 3 komponen yang dapat digunakan dalam analisis lebih lanjut. Ketiga komponen ini dapat menjelaskan 77,169% variabilitas data yang ada. *Loading factor* dari masing-masing komponen satu hingga tiga ada pada tabel 4. Nilai dari *loading factor* ini akan menjadi penetapan variabel mana yang akan masuk ke masing-masing komponen, berdasarkan nilai tertinggi

Tabel 3. Hasil *principal content analysis*

Importance of components									
	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9
Standard deviation	2,02676	1,26943	1,10728	0,95042	0,71253	0,50083	0,43680	0,33460	0,30029
Proportion of Variance	0,45642	0,17905	0,13623	0,10037	0,05641	0,02787	0,02120	0,01244	0,01002
Cumulative Proportion	0,45642	0,63547	0,77169	0,87206	0,92847	0,95634	0,97754	0,98998	1,00000

Tabel 4. *Loading factor* Komponen 1-3

Loadings:			
	Comp.1	Comp.2	Comp.3
murder	0,260	0,129	0,564
assault	0,411	-0,102	0,395
rape	0,184	-0,505	0,122
abduction	0,432	-0,191	-0,158
theftv	0,312	0,332	-0,476
theftwv	0,439	0,196	-0,169
drugs	-	0,583	-
fnc	0,324	0,364	0,306
capo	0,382	-0,248	-0,369

variabel tersebut. Kita dapat rumuskan anggota masing-masing komponen sebagai berikut:

- Komponen 1: assault, abduction, theftwv, capo
- Komponen 2: -rape, drugs, fnc
- Komponen 3: murder, -theftv

Setiap komponen dapat kita interpretasikan sesuai dengan anggota dan nilai serta tandanya, positif atau negatif. Komponen 1 menunjukkan daerah dengan tingkat kriminalitas serangan fisik, penculikan, pencurian tanpa kekerasan, dan kejahatan terhadap ketertiban umum yang tinggi. Komponen 2 menunjukkan daerah dengan kejahatan narkoba, penipuan dan korupsi yang relatif tinggi, namun kejahatan kesusilaan yang rendah. Komponen 3 menunjukkan daerah dengan tingkat pembunuhan yang tinggi, namun tingkat pencurian dengan kekerasan yang rendah.

#### 4.2. Validasi Internal dan Stabilitas Klaster

Validasi internal dan stabilitas dilakukan untuk menentukan jumlah optimal klaster dan metode *clustering* yang digunakan. Metode pengujian internal menggunakan ukuran Dunn, Silhouette, dan *Connectivity*. Hasil dari ketiga pengukuran tersebut dijabarkan pada tabel 5. Ukuran *Connectivity* dan Silhouette memiliki nilai optimal pada metode *Hierarchical* dengan jumlah klaster sebesar 2. Nilai kedua metode tersebut secara berturut-turut adalah 4,4829 dan 0,6064. Metode Dunn di sisi lain memiliki

Tabel 5. Validasi Internal masing-masing metode Clustering

Validation Measures:		2	3	4	5	6
hierarchical	Connectivity	4,48	7,29	11,14	12,64	18,00
	Dunn	0,46	0,54	0,56	0,56	0,37
	Silhouette	0,61	0,59	0,58	0,55	0,38
kmeans	Connectivity	8,96	10,04	11,14	18,04	19,54
	Dunn	0,36	0,47	0,56	0,19	0,24
	Silhouette	0,59	0,60	0,58	0,42	0,40
fanny	Connectivity	15,94	20,71	25,70	NA	NA
	Dunn	0,02	0,08	0,09	NA	NA
	Silhouette	0,20	0,28	0,25	NA	NA

Tabel 6. Rangkuman hasil Validasi Internal

Measure	Score	Method	Clusters
Connectivity	4,4829	hierarchical	2
Dunn	0,5629	hierarchical	4
Silhouette	0,6064	hierarchical	2

nilai optimal sebesar 0,5629 pada beberapa metode sekaligus, yaitu metode *Hierarchical* dengan 4 dan 5 klaster dan metode K-Means dengan 4 klaster. Nilai optimal keseluruhan ukuran ini dirangkum pada tabel 6.

Proses selanjutnya adalah melakukan validasi stabilitas masing-masing metode dengan ukuran APN, AD, ADM, dan FOM. Metode Fuzzy C-Means tidak dapat mencapai konvergensi pada validasi ini sehingga hasil validasi stabilitas ini hanya mencakup metode K-Means dan *Hierarchical*. Hasil masing-masing ukuran ini ada pada tabel 7. Ukuran AD dan FOM menyimpulkan bahwa metode K-Means dengan 6 klaster adalah yang paling optimal. Metode APN menyimpulkan bahwa metode K-Means dengan 2 metode adalah yang paling optimal. Metode ADM menyimpulkan bahwa metode *Hierarchical* dengan 5 klaster adalah yang paling optimal.

Tabel 7. Rangkuman hasil Validasi Stabilitas

Measures	Score	Clusters	Cluster
APN	0,0157	kmeans	2
AD	1,4794	kmeans	6
ADM	0,3194	hierarchical	5
FOM	1,0118	kmeans	6

#### 4.3. Jumlah Kluster Optimal

Proses validitas internal dan stabilitas telah memberikan jumlah optimal kluster dengan masing-masing pengukurannya. Selain itu, metode *Within-Cluster Sum of Squares* (WSS) juga akan digunakan dalam penentuan jumlah kluster. Hasil dari WSS ini dianalisis dengan *elbow method*, yaitu menentukan jumlah kluster berdasarkan tekukan kurva WSS. Kurva yang dihasilkan oleh WSS ditunjukkan pada gambar 2. Terdapat beberapa tekukan pada kurva ini, seperti pada kluster ke-2, 4, dan 6. Dengan hasil validasi internal dan stabilitas serta *elbow method* dari WSS, peneliti menggunakan sejumlah 2 kluster sebagai analisis *clustering*.

#### 4.4. K-Means Clustering

Proses analisis kluster metode K-Means dengan 2 kluster menghasilkan kelompok sebagai berikut:

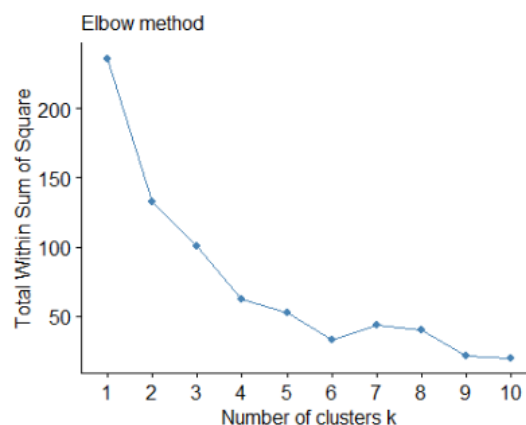
Kelompok 1: Sulawesi Utara, Gorontalo, Maluku, Papua Barat, Papua.

Kelompok 2: Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kepulauan Riau, Metro Jaya, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Sulawesi Barat, Maluku Utara.

Karakteristik tiap kelompok ditunjukkan pada tabel 8. Kelompok 1 memiliki karakteristik Komponen 1 dan 3 yang relatif tinggi dan komponen 2 yang relatif rendah. Kelompok 2 di lain sisi memiliki karakteristik berupa Komponen 1 dan 3 yang rendah dan komponen 2 yang tinggi.

#### 4.5. Hierarchical Clustering

Analisis kluster dengan metode *hierarchical* dilakukan dengan empat macam metode berdasarkan cara *linkage*: *complete linkage*, *average linkage*, *single*

Gambar 2. Kurva *Within-Cluster Sum of Squares*

Tabel 8. Karakteristik tiap kluster K-Means

Group	Comp.1	Comp.2	Comp.3
1	4.1180	-0,8414	0,3004
2	-0,7100	0,1451	-0,0518

Tabel 9. *Agglomerative Coefficient* Setiap metode *Hierarchical Clustering*

Method	Agglomerative Coef.
complete	0,8924
single	0,7892
average	0,8378
ward	0,9211

Tabel 10. Karakteristik tiap Kluster *Hierarchical* dengan metode Ward

Group	Comp.1	Comp.2	Comp.3
1	-0.7100	0.1451	-0.0518
2	4.1180	-0.8414	0.3004

*linkage*, dan *Ward's method*. Untuk menilai metode mana yang terbaik, dilakukan perbandingan masing-masing *agglomerative coefficient* yang mereka hasilkan. Nilai *agglomerative coefficient* ditunjukkan pada tabel 9. Metode Ward memiliki koefisien tertinggi, yaitu 0,9211, sehingga digunakan dalam analisis kluster *hierarchical*.

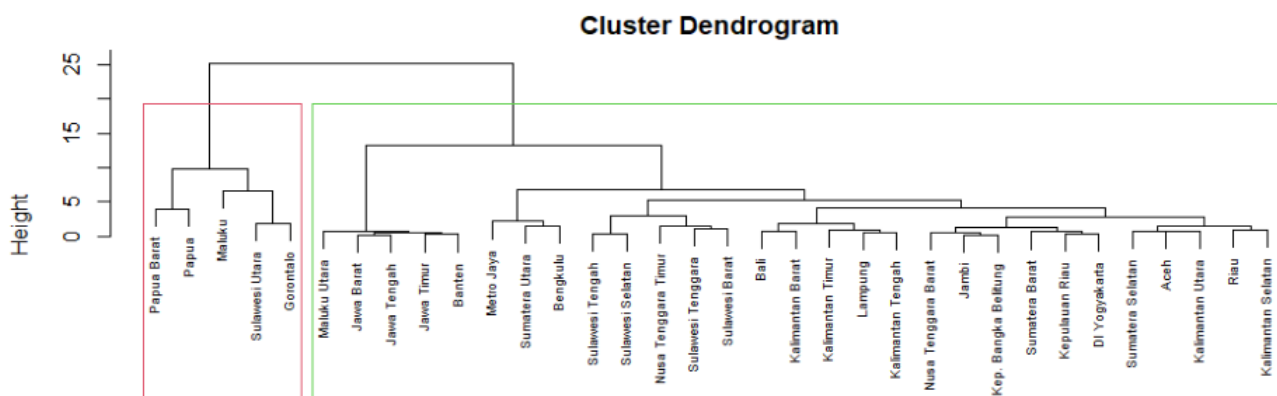
Karakteristik masing-masing kelompok dijelaskan pada tabel 10. Berikut adalah dua kelompok yang dihasilkan dengan *Hierarchical clustering*:

Kelompok 1: Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu,

Lampung, Kep. Bangka Belitung, Kepulauan Riau, Metro Jaya, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Sulawesi Barat, Maluku Utara.

Kelompok 2: Sulawesi Utara, Gorontalo, Maluku, Papua Barat, Papua.

Persebaran kelompok dapat dilihat pada dendrogram pada gambar 3.



Gambar 3. Dendrogram *Hierarchical Clustering* dengan metode Ward

#### 4.6. Fuzzy C-Means Clustering

Metode FCM menetapkan derajat keanggotaan pada masing-masing data. Derajat keanggotaan tersebut menjadi dasar penetapan klaster data. Tabel 11 menunjukkan derajat keanggotaan masing-masing data. Persentase paling tinggi menandakan klaster dimana data tersebut menjadi anggota.

Karakteristik masing-masing kelompok yang dihasilkan oleh FCM ditunjukkan pada tabel 12. Kelompok 1 memiliki karakteristik Komponen 1 dan 2 yang rendah, serta Komponen 3 yang tinggi secara relatif. Kelompok 2 memiliki karakteristik berupa Komponen 1 dan 2 yang tinggi, serta komponen 3 yang rendah secara relatif. Masing-masing anggota kelompok adalah sebagai berikut:

Kelompok 1: Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kepulauan Riau, Metro Jaya, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa

Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Sulawesi Barat, Maluku Utara.

Kelompok 2: Sulawesi Utara, Gorontalo, Maluku, Papua Barat, Papua.

#### 4.7. Perbandingan Hasil Tiap Metode

Hasil analisis *clustering* dengan metode yang berbeda memiliki hasil yang kurang lebih mirip. Terdapat perbedaan kodifikasi variabilitas komponen pada metode K-Means dengan *Hierarchical*.

Karakteristik Kelompok 1 dan Kelompok 2 pada kedua metode terbalik, tetapi memiliki nilai variabilitas komponen yang benar-benar sama (lihat tabel 8 dan 10). Oleh karena itu, hasil K-Means dengan *Hierarchical* sama, hanya saja penamaan kelompoknya terbalik. Selanjutnya, Kelompok K-Means akan dibalik agar seragam dengan hasil metode lain.

*Clustering* dengan metode FCM memiliki karakteristik yang berbeda dibanding dua metode lainnya. Contohnya, kelompok 1 pada metode FCM memiliki karakteristik berupa Komponen 1 dan 2 yang rendah, dan komponen 3 yang tinggi. Sementara itu, Kelompok 1 pada *Hierarchical Clustering* memiliki karakteristik Komponen 1 dan 3 yang rendah, dan Komponen 2 yang tinggi (lihat tabel 10 dan 12). Meskipun memiliki karakteristik yang berbeda, ketiga metode memiliki hasil klaster yang sama.

#### 4.8. Karakteristik Tiap Klaster dengan Variabel Kriminalitas

Tabel 11. Derajat keanggotaan Fuzzy C-Means (dalam %, dibulatkan)

Provinsi	Cluster 1	Cluster 2
Aceh	97	3
Sumatera Utara	58	42
Sumatera Barat	92	8
Riau	95	5
Jambi	99	1
Sumatera Selatan	90	10
Bengkulu	72	28
Lampung	98	2
Kep. Bangka Belitung	99	1
Kepulauan Riau	97	3
Metro Jaya	77	23
Jawa Barat	88	12
Jawa Tengah	89	11
DI Yogyakarta	98	2
Jawa Timur	91	9
Banten	90	10
Bali	95	5
Nusa Tenggara Barat	97	3
Nusa Tenggara Timur	84	16
Kalimantan Barat	97	3
Kalimantan Tengah	99	1
Kalimantan Selatan	89	11
Kalimantan Timur	95	5
Kalimantan Utara	93	7
Sulawesi Utara	30	70
Sulawesi Tengah	77	23
Sulawesi Selatan	75	25
Sulawesi Tenggara	92	8
Gorontalo	19	81
Sulawesi Barat	94	6
Maluku	31	69
Maluku Utara	90	10
Papua Barat	26	74
Papua	25	75

Tabel 12. Karakteristik tiap Klaster FCM

	Comp.1	Comp.2	Comp.3
Cluster 1	0,7312	0,1752	0,3570
Cluster 2	7,5418	0,2122	-2,7743

Tabel 13. Karakteristik tiap Klaster dengan Variabel Kriminalitas

Variabel	Cluster 1	Cluster 2
murder	0.0051	0.0128
assault	0.1795	0.6535
rape	0.0320	0.2129
abduction	0.0076	0.0681
theftv	0.0314	0.0718
theftwv	0.3574	0.8090
drugs	0.1864	0.0794
fnc	0.1606	0.3018
capo	0.0143	0.1235

Klaster yang telah terbentuk oleh ketiga metode disusun berdasarkan nilai komponen, yang merupakan gabungan dari variabel kriminalitas. Tabel 13 adalah karakteristik masing-masing klaster sesuai dengan variabel kriminalitas sebelum dijadikan komponen diukur dengan rata-rata setiap variabel.

Klaster 1 yang terdiri dari 29 provinsi dengan karakteristik berupa kriminalitas yang rendah, kecuali kriminalitas yang berkaitan dengan narkoba. Klaster 2 terdiri dari 5 Provinsi dengan tingkat kriminalitas relatif lebih tinggi, kecuali kriminalitas yang berkaitan dengan narkoba.

## 5. Kesimpulan

Penelitian ini menganalisis tingkat kriminalitas pada tiap provinsi/Polda berdasarkan jenisnya dengan mengelompokkannya dengan analisis *clustering*. Analisis *clustering* dengan metode K-Means, *Hierarchical*, dan Fuzzy C-Means menghasilkan dua klaster dengan anggota yang sama, didukung dengan hasil validasi internal dan stabilitas. Klaster 1 terdiri dari 29 provinsi dengan tingkat kejahatan relatif lebih rendah, kecuali kejahatan yang berkaitan dengan narkoba.



- <sup>1</sup> Institute of Economics and Peace. (2022). *Global Peace Index 2022: Measuring Peace in a Complex World*. Sydney: Tersedia pada: <http://visionofhumanity.org/resources>. Diakses pada 19 Desember 2022.
- <sup>2</sup> Numbeo.com. (2022). Crime Index by Country 2020-2022. Tersedia pada [https://www.numbeo.com/crime/rankings\\_by\\_country.jsp](https://www.numbeo.com/crime/rankings_by_country.jsp). Diakses pada 19 Desember 2022.
- <sup>3</sup> Badan Pusat Statistik. (2021). *Statistik Kriminal 2021*. Jakarta: Badan Pusat Statistik
- <sup>4</sup> Ibid, 114-117
- <sup>5</sup> Madhulatha, T. S. (2012). *An Overview on Clustering Method*. IOSR Journal of Engineering, 2(4), pp. 719
- <sup>6</sup> Jolliffe. I. T. & J. Cadima. (2016). *Principal Component Analysis: Review and Recent Developments*. Philosophical Transactions A, 374.
- <sup>7</sup> Pradaningtyas, D. A., T. Margawati, & J. T. Putro. (2022). *Disparities Among Districts in Central Java Province: Cluster Analysis Based on Several Well-Being Indicators*. Jurnal Ekonomi Pembangunan, 20(01), p. 38.
- <sup>8</sup> Amruthnath N. & T. Gupta. (2019). *Fault Diagnosis Using Clustering. What Statistical Test to Use for Hypothesis Testing?* Machine Learning and Applications: An International Journal, 6(1), p. 27
- <sup>9</sup> Madhulatha. *Overview on Clustering*, 721
- <sup>10</sup> Ferreira, L. & D. B. Hitchcock. (2009). *A Comparison of Hierarchical Methods for Clustering Functional Data*. Communications in Statistics – Simulation and Computation, p. 1927. DOI: 10.1080/03610910903168603
- <sup>11</sup> Deng, Songyin. (2019). *Clustering with Fuzzy C-Means and Common Challenges*. Journal of Physics: Conference Series, 1453 012137.
- <sup>12</sup> Liu, Y., Z. Li, H. Xiong, X. Gao, & J. Wu. (2010). *Understanding of Internal Clustering Validation Measures*. IEEE International Conference of Data Mining, p.911-912
- <sup>13</sup> Ibid.
- <sup>14</sup> Ibid.
- <sup>15</sup> Demir K. T. & Merenyi E.A. (2011). *A Validity Index for Prototype-Based Clustering of Data Sets with Complex Cluster Structures*. IEEE Transactions on systems, Man, and Cybernetics—Part B: Cybernetics, 41(4):1039-53.
- <sup>16</sup> Saputra, D. M., D. Saputra, & L. D. Oswari. (2019). *Effect of Distance Metrics in Determining K-Value in K-Means Clustering Using Elbow and Silhouette Method*. Advances in Intelligent Systems Research, 172, p. 343.