



# Mengungkap Lanskap Kejahatan Provinsi di Indonesia Tahun 2021: Analisis Perbandingan K-Means dan Logika Fuzzy

Afdatul Chofidah<sup>1</sup>, Setia Pramana<sup>2</sup>

<sup>1</sup>Program Studi Komputasi Statistik, Politeknik Statistika STIS

<sup>1</sup>[222111848@stis.ac.id](mailto:222111848@stis.ac.id)

<sup>2</sup>[setia.pramana@stis.ac.id](mailto:setia.pramana@stis.ac.id)

**Abstract:** *The crime rate in Indonesia has consistently decreased from 2017 to 2021. However, several provinces exhibit crime rates surpassing the national average due to localized factors. . This study aims to cluster Indonesian provinces based on crime indicators using the K-Means and Fuzzy clustering methods. Comparative analysis of cluster results relies on internal validity indices (connectivity, dunn, and silhouette) and stability metrics (APN, AD, ADM, and FOM). The findings indicate that the K-Means method excels in forming provincial clusters. K-Means achieves connectivity, dunn, and silhouette indices of 5.657, 0.913, and 0.549 respectively, while Fuzzy achieves 20.216, 0.183, and 0.223. Regarding stability, K-Means shows values of APN 0.011, AD 2.939, ADM 0.071, and FOM 0.676, whereas Fuzzy shows values of APN 0.011, AD 4.152, ADM 0.060, and FOM 0.916. The analysis resulted in the formation of 2 clusters, with Cluster 1 comprising 29 provinces and Cluster 2 comprising 4 provinces, each exhibiting distinct characteristics.*

**Keywords:** *clustering, K-Means, Fuzzy, crime, province*

**Abstrak:** Tingkat kejahatan Indonesia menunjukkan penurunan yang konsisten mulai tahun 2017-2021. Meskipun demikian, beberapa provinsi memiliki tingkat kejahatan yang melebihi tingkat kejahatan nasional. Hal tersebut disebabkan faktor lokal pada masing-masing provinsi. Penelitian ini bertujuan untuk mengelompokkan provinsi di Indonesia berdasarkan indikator-indikator kejahatan menggunakan metode K-Means dan Fuzzy dalam *clustering*. Perbandingan hasil klaster dilakukan berdasarkan indeks validitas internal (connectivity, dunn, dan silhouette) dan stabilitas (APN, AD, ADM, dan FOM). Hasil penelitian menunjukkan bahwa metode K-Means lebih unggul dalam membentuk klaster provinsi. Indeks connectivity, dunn, dan silhouette K-Means berturut-turut sebesar 5,657; 0,913; 0,549 sedangkan Fuzzy sebesar 20,216; 0,183; 0,223. Dalam hal stabilitas, K-Means memiliki nilai APN 0,011; AD 2,939; ADM 0,071; dan FOM 0,676, sementara Fuzzy memiliki nilai APN 0,011; AD 4,152; ADM 0,060; dan FOM 0,916. Jumlah klaster yang terbentuk adalah 2, dengan klaster 1 terdiri dari 29 provinsi dan klaster 2 terdiri dari 4 provinsi, masing-masing dengan karakteristik yang berbeda.

**Kata kunci:** *clustering, K-Means, Fuzzy, kejahatan, provinsi*

## I. PENDAHULUAN

Kriminalitas atau tindak kejahatan adalah suatu tindakan yang melanggar undang-undang, hukum, nilai, dan norma-norma yang berlaku di masyarakat[1]. Tindak kejahatan menjadi ancaman utama bagi kehidupan manusia[2]. Tindak kejahatan ini tidak hanya membahayakan, tetapi juga merugikan keselamatan serta jiwa seseorang.

Berdasarkan gambar 1, tingkat kriminalitas Indonesia menunjukkan tren yang konsisten menurun. Tingkat kejahatan meruapaka angka yang menunjukkan seberapa rentan suatu lokasi terhadap kejahatan dalam jangka waktu tertentu. Semakin tinggi tingkat kejahatan, semakin besar kerentanan terhadap kejahatan di daerah tersebut[3]. Pada tahun 2021, tingkat kejahatan Indonesia mencapai 90 jiwa. Artinya, dalam 100.000 penduduk, terdapat 90 orang yang mengalami atau menjadi korban kejahatan. Penurunan tersebut menunjukkan bahwa semakin baiknya pemerintah dalam menangani berbagai tindak kejahatan. Meskipun begitu, masih terdapat provinsi dengan tingkat kejahatan yang melebihi tingkat kejahatan nasional. Pada tahun 2021, provinsi dengan tingkat kejahatan tertinggi adalah Papua Barat yaitu sebesar 289 jiwa, diikuti dengan DKI Jakarta dan Sumatera Utara. Sebaliknya, provinsi dengan tingkat kejahatan paling rendah adalah Jawa Barat sebesar 15 jiwa. Hal ini membuktikan bahwa ada faktor-faktor lokal yang mempengaruhi tingkat kejahatan masing-masing provinsi.



**Gambar 1.** Grafik tingkat kejahatan Indonesia dan tiga provinsi tertinggi tahun 2017-2021

Salah satu pemicu terjadinya tindak kejahatan adalah kepadatan penduduk. Daerah dengan populasi yang padat menyebabkan tingginya persaingan antraindividu untuk bertahan hidup. Peningkatan penduduk yang tidak diimbangi dengan persebaran yang merata dapat menyebabkan berkurangnya lapangan pekerjaan[4]. Akibatnya, jumlah penangguran akan terus meningkat dan ketimpangan pendapatan antara yang bekerja semakin tinggi[4]. Beberapa individu merasa terdesak akan melakukan tindak kejahatan seperti pencurian, perampokan, atau penipuan demi memenuhi kebutuhan mereka. Selain itu, salah satu faktor yang signifikan mempengaruhi tindak kejahatan adalah kemiskinan. Tekanan ekonomi yang tinggi dapat menjadi dorongan bagi individu untuk mencari penghasilan tambahan dengan cara yang illegal[5].

Faktor lain pemicu tindak kejahatan adalah tingkat pendidikan. Semakin rendah tingkat pendidikan seseorang, kecenderungan keterampilan yang dimiliki juga semakin rendah[6]. Individu yang hanya lulus pendidikan dasar memiliki lebih banyak waktu luang dibandingkan dengan individu yang menyelesaikan pendidikan hingga SMA dan perguruan tinggi. Waktu luang yang lebih banyak dapat disalahgunakan bagi individu untuk melakukan tindak kriminal.

Berdasarkan faktor-faktor tersebut, pengelompokan provinsi berdasarkan indikator kejahatan penting untuk mengidentifikasi karakteristik kejahatan di berbagai wilayah. Pemerintah dapat merancang kebijakan yang efektif sesuai dengan karakteristik wilayah tersebut untuk terus menurunkan tingkat kejahatan di Indonesia. Banyak penelitian terkait pengelompokan provinsi/kota berdasarkan tingkat kejahatannya. Metode-metode yang digunakan pada penelitian tersebut diantaranya hierarchical clustering[7], fuzzy c-means[8], dan hierarchical clustering dan k-means[9]. Namun, penelitian-penelitian tersebut umumnya hanya menggunakan variabel kejahatan dan tidak mempertimbangkan faktor-faktor penyebab kejahatan dalam pembentukan kluster. Penelitian ini berupaya mengisi gap tersebut dengan metode yang berbeda dari penelitian sebelumnya.

Secara umum, metode dalam clustering dibedakan menjadi dua yaitu *hard clustering* dan *soft clustering*. Penelitian ini menggunakan serta membandingkan metode *hard clustering* dan *soft clustering* dalam analisis klasternya. Pendekatan yang digunakan adalah K-Means dan Fuzzy C-Means. Dengan kedua pendekatan tersebut, penelitian ini juga memperhitungkan faktor penyebab kejahatan. Selain itu penelitian bertujuan untuk melihat gambaran provinsi-provinsi di Indonesia berdasarkan lanskap atau indikator kejahatan dengan mengetahui karakteristik tiap kelompoknya.



## II. METODOLOGI

### 2.1. Data dan Sumber Data

Data yang digunakan pada penelitian ini merupakan data sekunder dari tabel Badan Pusat Statistik dan Publikasi “Statistik Kriminal 2023”. Penelitian menggunakan data tingkat provinsi tahun 2021. Adapun dataset terdiri dari 16 variabel/fitur yang tersedia pada Tabel 1.

Tabel 1. Rincian Variabel Penelitian

Kode	Nama Variabel	Deskripsi
kepadatan	Kepadatan penduduk	Banyaknya penduduk per satuan luas(jiwa/km <sup>2</sup> )
tpt	Tingkat pengangguran terbuka	Persentase jumlah pengangguran terhadap jumlah angkatan kerja(%)
rls	Rata-rata lama sekolah	Rata-rata lama sekolah menurut metode baru pengukuran ipm (tahun)
miskin	Persentase penduduk miskin	Persentase penduduk miskin yang berada di bawah garis kemiskinan(%)
resiko	Risiko penduduk terkena kejahatan	Risiko penduduk terkena kejahatan per 100.000 penduduk
persen_selesai	Persentase penyelesaian kejahatan	Persentase penyelesaian kejahatan(%)
jht_nyawa	Kejahatan terhadap nyawa	Jumlah kejahatan terhadap nyawa mencakup pembunuhan
jht_fisik	Kejahatan terhadap fisik/badan	Jumlah kejahatan terhadap fisik yang mencakup penganiayaan berat, penganiayaan ringan, dan kekerasan dalam rumah tangga.
jht_asusila	Kejahatan terhadap kesusilaan	Jumlah kejahatan asusila mencakup perkosaan dan pencabulan
jht_kemerdekaan	Kejahatan terhadap kemerdekaan orang	Jumlah kejahatan terhadap kemerdekaan orang terdiri dari jenis kejahatan penculikan dan mempekerjakan anak di bawah umur
jht_hak_milik_keras	Kejahatan terhadap hak milik/barang dengan penggunaan kekerasan	Jumlah kejahatan terhadap hak milik/barang dengan penggunaan kekerasan terdiri dari pencurian dengan kekerasan menggunakan senjata api (senpi), dan pencurian dengan kekerasan menggunakan senjata tajam (sajam).
jht_milik	Kejahatan terhadap hak milik/barang	Jumlah kejahatan terhadap hak milik/barang tanpa penggunaan kekerasan terdiri dari pencurian, pencurian dengan pemberatan, pencurian kendaraan bermotor, pengrusakan/penghancuran barang, pembakaran dengan sengaja, dan penadahan
jht_narkotika	Kejahatan terkait narkotika	Jumlah kejahatan terkait narkotika mencakup kejahatan narkotika (narkoba) dan psikotropika.
jht_dana	Kejahatan terkait penipuan, penggelapan, dan korupsi	Jumlah kejahatan ini mencakup kejahatan penipuan, penggelapan, dan korupsi.
jht_tertib	Kejahatan terhadap ketertiban umum	Jumlah kejahatan terhadap ketertiban umum

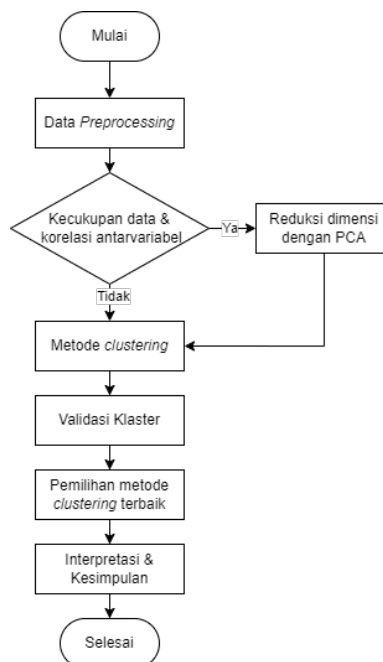
### 2.2. Data Preprocessing

Data preprocessing merupakan suatu proses untuk mengoreksi data mulai dari noisy data, missing value, hingga sebaran data sehingga menghasilkan data yang berkualitas dan layak untuk diolah[14].

Pada penelitian ini, tahap preprocessing yang dilakukan adalah eksplorasi noisy data dan standarisasi data.

### 2.3. Clustering

*Clustering* atau analisis kluster adalah metode eksplorasi data untuk mendapatkan karakteristik data pada data yang tidak memiliki informasi apriori berupa label dan mekanisme pengelompokan (*unsupervised learning*)[10]. *Clustering* memanfaatkan informasi, karakteristik, dan fitur dari tiap objek dalam mengelompokkan objek. Objek dengan karakteristik yang serupa dikelompokkan ke dalam satu kluster sedangkan objek yang karakteristiknya berbeda ditempatkan pada kluster yang lain. Hal ini bertujuan untuk memaksimalkan kemiripan antarobjek dalam kluster dan meminimalkan kemiripan antarkluster yang berbeda[11]. Alur *clustering* dalam penelitian ini secara singkat dijelaskan pada Gambar 2.



Gambar 2. Diagram alir penelitian

Penelitian ini adalah menggunakan K-Means dan Fuzzy C Means dalam analisis klasternya.

#### 1. *K-Means Clustering*

K-Means termasuk dalam metode *hard-clustering* dengan setiap titik berada tepat pada satu kluster. Setiap kluster akan direpresentasikan oleh sebuah titik pusat (centroid) kemudian setiap objek akan dikelompokkan ke kluster dengan titik pusat terdekat[11]. Karena itu, kelebihan metode ini adalah sangat sederhana .

#### 2. *Fuzzy C-Means Clustering*

Fuzzy C-Means (FCM) merupakan salah satu metode *soft-clustering* yang paling populer. Fuzzy C-Means memungkinkan suatu objek menjadi anggota dari satu atau lebih kluster yang terbentuk dengan derajat keanggotaan antara 0 hingga 1[12]. Derajat keanggotaan ini yang akan menentukan tingkat keberadaan objek dalam suatu kluster. Kelebihan dari Fuzzy C-Means adalah kemampuannya yang robust dalam meminimumkan fungsi objektif. Namun, kekurangannya, metode ini mudah terjebak pada optimum lokal dan sensitif terhadap *noise* [13].



#### 2.4. Reduksi Dimensi

Dengan mengabaikan variabel ‘provinsi’ yang mewakili identitas wilayah, jumlah variabel pada penelitian ini adalah 16 variabel sehingga dapat disebut berdimensi tinggi.

##### **Kaiser-Meyer-Olkin (KMO) dan Bartlett’s Test**

KMO merupakan uji untuk menentukan kecukupan sampel. Nilai KMO berada pada rentang 0 hingga 1. Nilai yang dihasilkan oleh KMO *Measure of Sampling Adequacy* harus melebihi 0,5 untuk melanjutkan PCA. Selain itu, ada uji *Bartlett’s Test of Sphericity* yang dilakukan untuk memastikan bahwa variabel dalam set data tersebut tidak berkorelasi satu sama lain. Signifikansi dalam uji Bartlett’s ini harus bernilai kurang dari 0,05 agar dapat dilakukan PCA[14].

##### **Principal Component Analysis (PCA)**

PCA merupakan salah satu metode untuk mereduksi dimensi. Metode ini dengan membentuk satu set komponen baru yang merupakan kombinasi linier dari set variabel asal dengan tetap mempertahankan total keragaman variabel asal sehingga informasi awal tidak berkurang. Set komponen baru yang telah terbentuk tersebut dapat menggantikan variabel dan hubungan antar komponennya sudah saling bebas[11].

#### 2.5. Pemilihan Jumlah Kluster

Dalam menentukan jumlah kluster optimal, penelitian ini menggunakan tiga metode:

1. Metode elbow

Metode ini memilih jumlah kluster terbaik melalui sudut yang terbentuk. Jumlah kluster optimal ditentukan berdasarkan sudut atau penurunan yang paling signifikan. Semakin tinggi penurunan, jumlah kluster tersebut dianggap yang semakin optimal[15].

2. Metode silhouette

Metode ini menghitung rata-rata jarak antara satu objek dengan seluruh objek lain terdapat pada kluster yang sama, serta dengan objek di kluster lainnya. Jika nilainya semakin mendekati satu, semakin optimal jumlah kluster tersebut[13].

3. Metode *gap statistics*

Metode ini menentukan jumlah kluster optimum melalui nilai gap tertinggi. Jika nilai gap terus meningkat, jumlah kluster optimum ditentukan melalui nilai yang menunjukkan kenaikan gap paling kecil [13].

#### 2.6. Validasi Kluster

Validasi kluster dapat dilakukan melalui validitas internal dan stabilitas. Terdapat dua kriteria validasi internal yaitu *compactness* dan *separation*. *Compactness* mengukur kemiripan objek-objek dalam satu kluster sedangkan *separation* mengukur perbedaan objek-objek antarkluster yang berbeda[14]. Indeks validitas internal yang digunakan pada penelitian ini adalah indeks *Connectivity*, *Dunn*, dan *Silhouette*. *Connectivity* berada pada rentang hingga tak hingga. Kluster semakin baik jika *connectivity*nya seminimum mungkin[13]. Indeks *Dunn* menghitung nilai minimum dari rasio antara nilai ketidaksamaan dua kluster sebagai ukuran *separation* dan nilai maksimum dari diameter kluster sebagai ukuran *compactness*. Semakin besar indeks *dunn* menunjukkan bahwa jumlah kluster semakin optimal[16]. Sementara indeks *silhouette* menghitung rata-rata perbedaan jarak rata-rata dalam kluster dan jarak minimum antarkluster. Kluster makin baik jika indeks *silhouette* mendekati satu[13].

Selain validasi internal, validasi stabilitas juga digunakan untuk membandingkan hasil analisis kluster. Pengukuran ini mengevaluasi kekonsistenan kluster yang terbentuk dengan membandingkannya dengan kluster baru, kolom per kolom[14]. Ukuran yang digunakan pada validasi stabilitas penelitian ini adalah *Average Distance Between Means (ADM)*, *Average Proporsion of Non-overlap (APN)*,



Average *Distance* (AD), dan *Figure of Merit* (FOM). Semakin kecil nilai dari ukuran tersebut, semakin baik cluster yang dibentuk.

### III. HASIL DAN PEMBAHASAN

#### 3.1. Analisis Deskriptif

**Tabel 2.** Ringkasan Statistik Variabel Penelitian

Kode	Min	Q1	Q2	Mean	Q3	Max
kepadatan	9	54,25	103,50	744,26	261	15.978
tpt	3,01	4,45	5,05	5,49	6,32	9,91
rls	7,05	8,54	9,22	9,26	9,70	11,20
miskin	4,56	6,41	8,51	10,43	12,64	27,38
resiko	15	91	116,5	130	168,2	289
persen_selesai	32,83	57,24	67,03	70,03	78,25	143,65
jht_nyawa	0	9,25	20,50	27,26	42	96
jht_fisik	24	311,2	462	826,2	1089,8	4287
jht_asusila	20	62,75	114	173,68	231,75	904
jht_kemerdekaan	0	3	17,5	48,68	66,75	375
jht_hak_milik_keras	1	31,75	75,50	142,29	194,75	564
jht_milik	92	810	1184	2040	2562	12433
jht_narkotika	2	209	410,4	2086,9	1383,5	5949
jht_dana	40	314,5	566,5	1032,1	1003,8	5439
jht_tertib	0	0,25	4	74,47	87	616

Tabel 2 menyajikan ukuran pemusatan serta nilai minimum maksimum variabel. Ukuran pemusatan atau ukuran Lokasi menunjukkan bagaimana distribusi dari suatu data[17]. Ukuran pemusatan yang digunakan adalah rata-rata, median, dan kuantil. Berdasarkan tabel 2, terlihat bahwa ada beberapa variabel yang nilai minimumnya 0. Variabel-variabel tersebut adalah kejahatan terhadap nyawa, kejahatan terhadap kemerdekaan orang, dan kejahatan terhadap ketertiban umum. Selain itu, terlihat bahwa jarak antara nilai kuartil atas (Q3) dengan nilai maksimumnya sangat jauh pada variabel kepadatan. Dengan demikian, perlu dilakukan tahap preprocessing data untuk mengeksplorasi data dan memeriksa anomali pada data.

#### 3.2. Data Preprocessing

##### Pemeriksaan *noisy data*

Setelah dilakukan pengolahan data, tidak ada variabel yang memiliki missing value. Semua variabel lengkap sehingga dilanjutkan untuk memeriksa nilai minimum 0 pada variabel jumlah kejahatan. Identifikasi provinsi yang memiliki nilai 0 tersebut disajikan pada Tabel 4.

**Tabel 3.** Identifikasi Nilai Minimum 0

Variabel	Provinsi
jht_nyawa	Maluku Utara
jht_kemerdekaan	Sulawesi tengah
jht_tertib	Lampung, Riau, Nusa Tenggara Barat, Kepulauan Bangka Belitung, Banten, Kalimantan Utara, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Barat

Statistik Kriminal melakukan penghitungan jumlah kejahatan menggunakan pendekatan jumlah pelaporan[18]. Berdasarkan keterangan pada publikasi Statistik Kriminal, nilai 0 artinya memang tidak ada kasus kejahatan pada daerah tersebut sehingga nilai tersebut dianggap benar. Selanjutnya adalah pengecekan *outlier* pada data.



Berdasarkan tabel ringkasan statistik variabel, variabel kepadatan memiliki jarak yang sangat jauh antara nilai kuartil atas dan nilai maksimumnya. Setelah diidentifikasi, provinsi dengan nilai kepadatan tertinggi tersebut adalah DKI Jakarta yaitu sebesar 15.978. Nilai tersebut dapat dikategorikan sebagai *outlier*. Hampir seluruh variabel memiliki *outlier*. Hanya variabel kejahatan terhadap nyawa yang tidak memiliki *outlier*. Dalam hal ini, *outlier* tidak dihapus untuk menjaga informasi penting dan kondisi spesifik di provinsi tertentu. Misalnya pada provinsi DKI Jakarta, nilai kepadatannya dianggap valid mengingat DKI Jakarta merupakan daerah ibukota dan kepadatannya lebih tinggi dibandingkan provinsi lainnya.

### Standarisasi

Standarisasi dilakukan pada semua variabel numerik. Standarisasi dilakukan menggunakan metode Z-score untuk menghasilkan nilai normal baku.

### 3.3.Reduksi Dimensi

Sebelum dilakukan clustering, perlu dilakukan pengecekan asumsi bahwa data cukup secara statistik dan tidak ada multikolinieritas pada data.

1. Uji KMO dan Bartlett

**Tabel 4.** Nilai KMO dan Uji Bartlett

Uji	Nilai	
KMO	0.68	
Bartlett	Approx. Chi-Square	432,88
	Degrees of Freedom	105
	P-Value	$2,22 \times 10^{-16}$

Berdasarkan tabel 4, nilai KMO lebih besar dari 0,6. Artinya, set data cukup secara statistik. Selain itu, *p-value* uji Bartlett kurang dari  $\alpha = 0,05$ . Hal ini berarti terdapat hubungan atau korelasi nyata antarvariabel. Hasil dari kedua uji tersebut menunjukkan bahwa perlu dilakukan reduksi dimensi pada dataset.

2. *Principal Component Analysis* (PCA)

Dalam penerapan PCA, nilai eigen yang melebihi 1 dipakai sebagai kriteria untuk menentukan jumlah komponen utama yang digunakan.

**Tabel 5.** Nilai Eigen dan Proporsi Varians PC

Kode	Nilai Eigen	Persentase Kumulatif Varians
PC1	5,8532	0,3902
PC2	2,8129	0,5777
PC3	1,7114	0,6918
PC4	1,0614	0,7626
PC5	0,9295	0,8245

Berdasarkan tabel 5, empat komponen secara bersama-sama sudah dapat menjelaskan 76,26% dari total seluruh varians variabel asal.

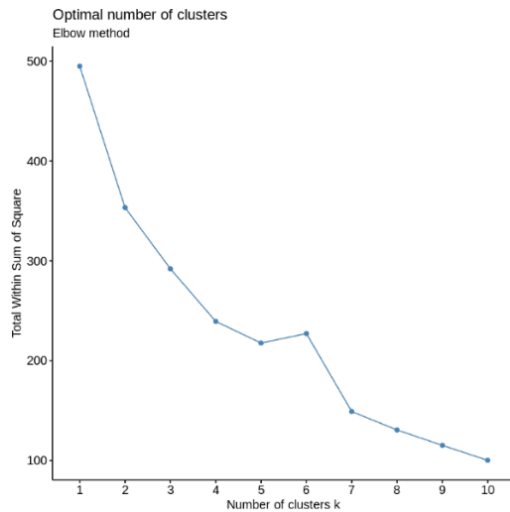
### 3.4.Pembentukan klaster

1. *K-Means Clustering*

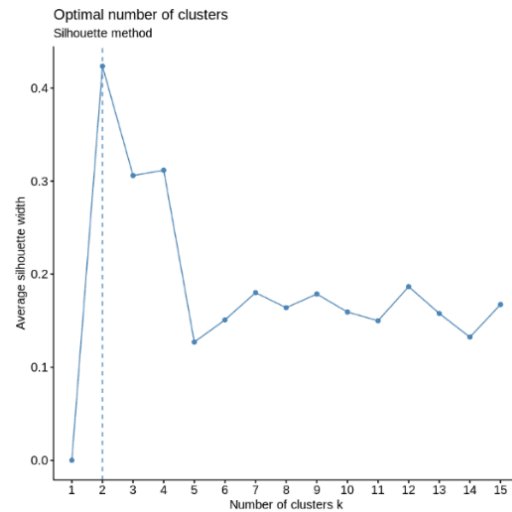
Berdasarkan *elbow plot* terlihat, terlihat garis yang membentuk siku (elbow) atau mengalami patahan adalah saat  $k = 4$ . Jumlah klaster optimal untuk metode K-Means berada dalam rentang 2 hingga 4 klaster. Berdasarkan metode silhouette, nilai  $k$  yang optimal terlihat pada garis paling tinggi dan berada di garis putus-putus. Grafik menunjukkan bahwa titik kedua memiliki nilai tertinggi yang menunjukkan bahwa nilai  $k$  yang optimal adalah 2. Sementara, pada metode *gap*



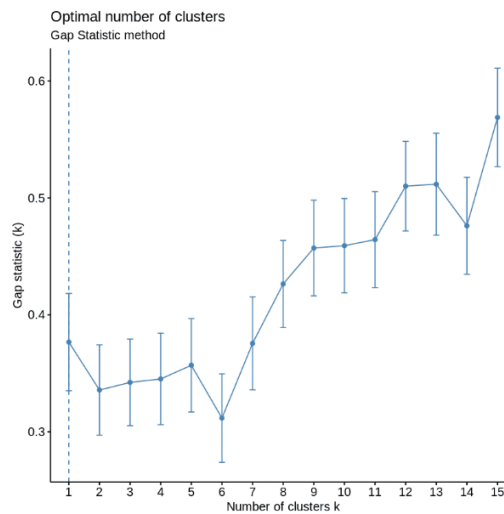
*statistic*, titik awal saat terjadi peningkatan yang pertama kali terletak pada  $k=2$ . Berdasarkan ketiga plot tersebut, jumlah kluster optimal untuk K-Means adalah 2 kluster.



**Gambar 3a.**



**Gambar 3b.**

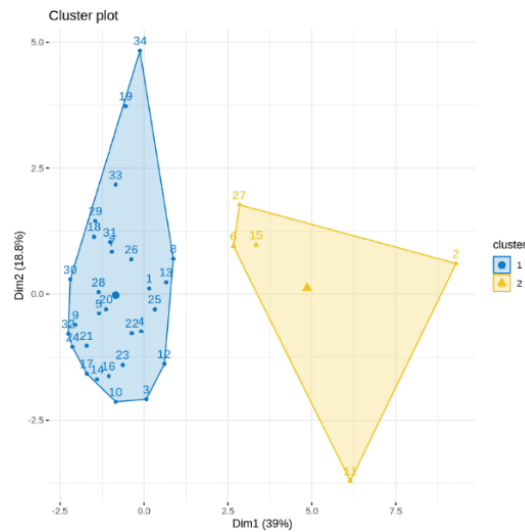


**Gambar 3c.**

**Gambar (3a)** *Elbow plot K-Means clustering* (3b) *Grafik shillouette K-Means clustering*  
(3c) *Grafik gap statistics K-Means clustering*

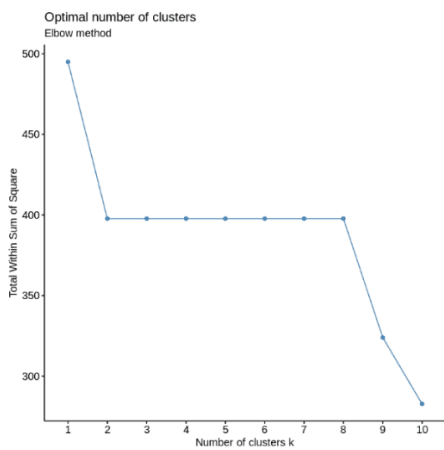
Gambar 4 menunjukkan hasil pengelompokkan berdasarkan K-Means. Pengelompokkan menggunakan K-Means menghasilkan 2 kluster dengan kluster 1 beranggotakan 29 provinsi dan kluster 2 beranggotakan 5 provinsi.



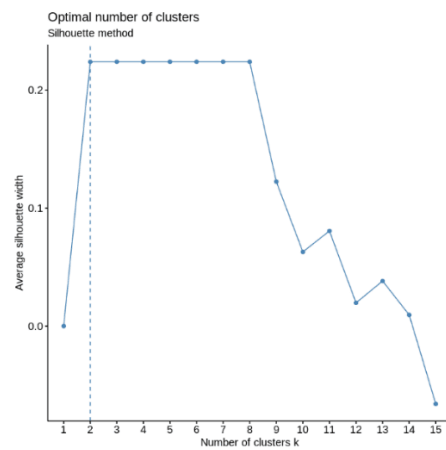


**Gambar 4.** Hasil clustering menggunakan K-Means

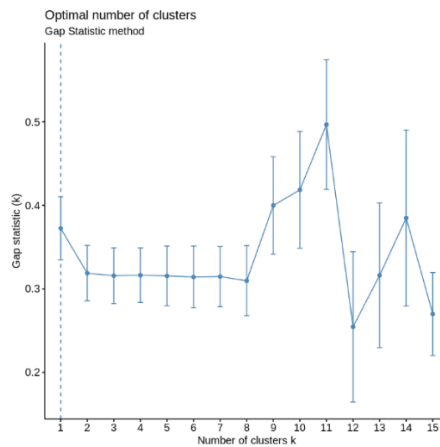
## 2. Fuzzy C-Means Clustering



**Gambar 5a.**



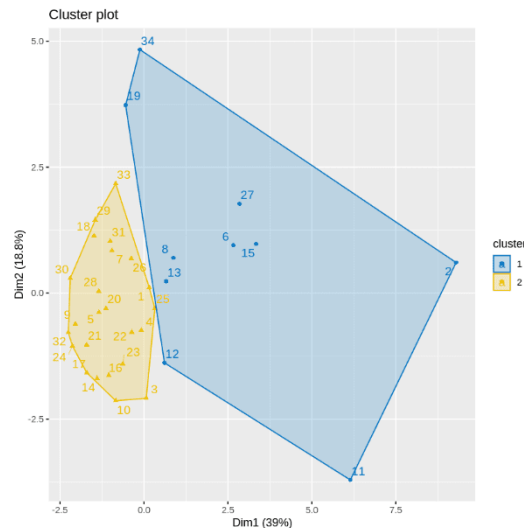
**Gambar 5b.**



**Gambar 5b.**

**Gambar (5a)** Elbow plot Fuzzy C-Means clustering (5b) Grafik silhouette Fuzzy C-Means clustering (5c) Grafik gap statistics Fuzzy C-Means clustering

Berdasarkan *elbow plot* terlihat, patahan atau siku (elbow) terlihat pada saat  $k=2$ . Metode silhouette juga menunjukkan titik yang paling tinggi dan berada di garis putus-putus adalah 2. Pada metode *gap statistic*, peningkatan pertama kali terjadi pada  $k=2$ . Pada metode fuzzy, jumlah kluster optimal juga 2 kluster.



**Gambar 6.** Hasil clustering menggunakan Fuzzy C-Means

Hasil pengelompokan berdasarkan Fuzzy-C-Means dapat dilihat pada gambar 6. Pengelompokan menggunakan Fuzzy menghasilkan 2 kluster dengan kluster 1 beranggotakan 9 provinsi dan kluster 2 beranggotakan 24 provinsi.

### 3. Pemilihan Kluster Terbaik

**Tabel 6.** Ringkasan Validitas Internal Kluster

Indeks	Skor K-Means	Skor Fuzzy
Connectivity	5,6579	20,2159
Dunn	0,9126	0,1831
Silhouette	0,5495	0,2239

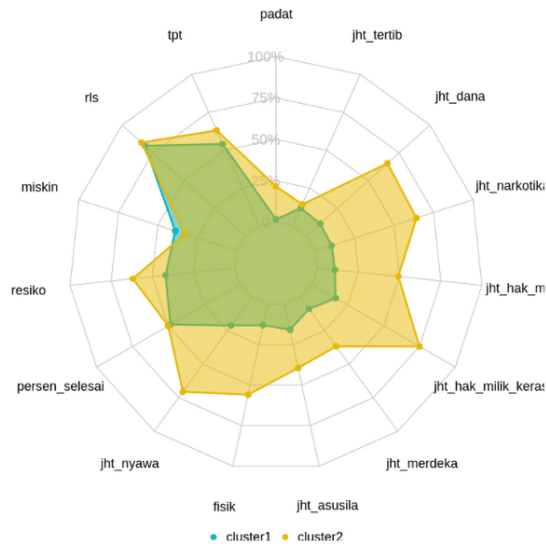
**Tabel 7.** Ringkasan Validitas Stabilitas K-Means

Kriteria	Skor K-Means	Skor Fuzzy
APN	0,0107	0,0111
AD	2,9381	4,1521
ADM	0,0711	0,0597
FOM	0,6756	0,9158

Berdasarkan tabel 6, indeks connectivity K-Means lebih kecil dari Fuzzy C-Means sedangkan indeks dunn dan silhouette-nya lebih tinggi. Hal ini menunjukkan bahwa K-Means lebih unggul dalam membentuk kluster. Selain itu 3 dari 4 kriteria validitas stabilitas, nilai K-Means lebih rendah dari Fuzzy. Dapat disimpulkan, metode K-Means lebih unggul untuk mengelompokkan provinsi di Indonesia berdasarkan indikator kejahatan.

### 4. Analisis Kluster Terpilih

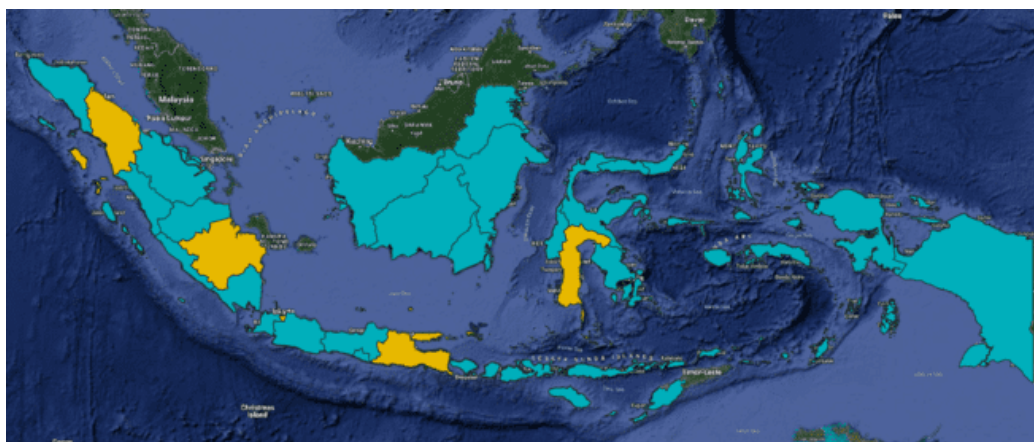
Setelah diperoleh kluster terbaik dan valid, yaitu K-Means, karakteristik tiap kluster dapat diketahui melalui rata-rata tiap indikator. Dalam hal ini rata-rata tiap indikator tiap kluster disajikan dalam bentuk *radar chart*.



**Gambar 7.** Radar chart variabel kluster

Berdasarkan radar chart pada Gambar 7, kluster 1 adalah kluster dengan indikator kejahatan yang rendah. Kluster ini memiliki kepadatan penduduk yang rendah dan resiko penduduk terkena kejahatan yang rendah. Jumlah kejahatan pada kluster ini juga rendah ditunjukkan dengan titik biru yang mendekati 0% dibandingkan dengan titik kuning.

Sementara itu, kluster 2 secara umum, jumlah kejahatannya tinggi. Hal ini terlihat dari titik kuning yang mendekati 100%. Populasinya padat dan resiko penduduk terkena kejahatan tinggi. Indikator-indikator lain seperti tingkat pengangguran terbuka, rata-rata lama sekolah, persentase penduduk miskin, dan persentase penyelesaian kejahatan pada kluster 1 lebih baik daripada kluster 2 walaupun perbedaannya tidak signifikan. Berdasarkan indikator-indikator tersebut, daerah-daerah yang berada di kluster pertama cenderung lebih aman dari tindak kejahatan dibandingkan dengan daerah pada kluster kedua.



**Gambar 8.** Peta Indonesia hasil clustering

Berdasarkan gambar 8, terdapat 5 provinsi yang dikategorikan kluster 2 atau provinsi dengan indikator kejahatan tinggi. Provinsi tersebut adalah Sumatera Selatan, Sumatera Utara, Sulawesi Selatan, DKI Jakarta, dan Jawa Timur.



#### IV. KESIMPULAN

Berdasarkan hasil dan pembahasan, ukuran optimal kluster adalah 2. Pemilihan jumlah kluster optimal dilakukan menggunakan *elbow plot*, grafik silhouette, dan *gap statistics*. Selain itu, metode K-Means lebih baik dalam mengelompokkan provinsi di Indonesia berdasarkan indikator kejahatan. Hal tersebut dapat dilihat dari indeks validitas internal dan stabilitasnya yang lebih unggul dari Fuzzy. Indeks validitas internal K-Means yaitu indeks connectivity 5,657; indeks dunn 0,913; dan indeks silhouette 0,549. Sementara, validitas stabilitasnya adalah APN 0,011; AD 2,939; ADM 0,071; dan FOM 0,676.

Secara umum, kluster yang terbentuk dari K-Means memiliki karakteristik yang berbeda. Kluster pertama yang beranggotakan 29 provinsi cenderung memiliki indikator kejahatan yang lebih baik dengan jumlah kejahatan rendah. Sementara, kluster kedua yang beranggotakan 5 provinsi memiliki indikator kejahatan lebih buruk dan jumlah kejahatan lebih tinggi.

#### UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Politeknik Statistika STIS yang telah memberikan dukungan dana penelitian dalam menyelesaikan kajian ini.

#### REFERENSI

1. K. Kartono, *Patologi Sosial*. 1981.
2. A. D. Putra, G. S. Martha, M. Fikram, and R. J. Yuhan, “Faktor-Faktor yang Memengaruhi Tingkat Kriminalitas di Indonesia Tahun 2018,” *Indones. J. Appl. Stat.*, vol. 3, no. 2, p. 123, 2021, doi: 10.13057/ijas.v3i2.41917.
3. M. A. Juniar, A. Fania, D. Ulya, R. Ramadhan, and N. Chamidah, “MODELLING CRIME RATES IN INDONESIA USING TRUNCATED SPLINE ESTIMATOR,” vol. 18, no. 2, pp. 1201–1216, 2024.
4. R. M. Sabiq and N. Nurwati, “Pengaruh Kepadatan Penduduk Terhadap Tindakan Kriminal,” *J. Kolaborasi Resolusi Konflik*, vol. 3, no. 2, p. 161, 2021, doi: 10.24198/jkrk.v3i2.35149.
5. R. Saputra, “Analisis Tingkat Pendidikan, Kemiskinan dan Pengangguran Terhadap Kriminalitas di Bekasi,” *J. Ilmu Hukum, Hum. dan Polit.*, vol. 3, no. 4, pp. 159–163, 2023, doi: 10.38035/jihhp.v3i4.1677.
6. L. Lochner, “Individual Perceptions of the Criminal Justice System,” *AEA J. Artic. Full-Text Access*, vol. 97, 2007.
7. R. Gustriansyah, J. Alie, and N. Suhandi, “Hierarchical clustering for crime rate mapping in Indonesia,” *Ilk. J. Ilm.*, vol. 14, no. 3, pp. 275–283, 2022, doi: 10.33096/ilkom.v14i3.1135.275-283.
8. J. Inayah, D. A. S. N. Maghfiroh, and D. C. R. Novitasari, “Clustering Daerah Rawan Kriminalitas Menggunakan Algoritma Fuzzy C-Means,” *J. Ilm. Inform. Komput.*, vol. 27, no. 2, pp. 95–106, 2022, doi: 10.35760/ik.2022.v27i2.6019.
9. M. F. Mardianto, Kuzair, T. Yulianto, R. Amalia, and Faisol, “Pengelompokan Optimal Kabupaten dan Kota Rawan Kriminalitas di Jawa Timur dengan Metode Analisis Kluster Terbaik,” *Zeta*, vol. 1, 2015.
10. A. Asra, A. P. Utomo, M. Asikin, and N. H. Puspongoro, *Analisis multivariabel suatu pengantar*. Bogor: IN MEDIA, 2017.
11. S. Pramana, B. Yuniarto, S. Mariyah, I. Santoso, and R. Nooraeni, *Data mining dengan R konsep setara implementasi*. Bogor: IN MEDIA, 2018.
12. D. L. Rahakbauw, V. Y. I. Ilwaru, and M. H. Hahury, “Implementasi Fuzzy C-Means Clustering Dalam Penentuan Beasiswa,” *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 11, no. 1, pp. 1–12, 2017, doi: 10.30598/barekengvol11iss1pp1-12.
13. N. Thamrin and A. W. Wijayanto, “Comparison of Soft and Hard Clustering: A Case Study on Welfare Level in Cities on Java Island,” *Indones. J. Stat. Its Appl.*, vol. 5, no. 1, pp. 141–160, 2021, doi: 10.29244/ijsa.v5i1p141-160.
14. A. Az-Zahra and A. W. Wijayanto, “Tinjauan Kesejahteraan di Daerah Perbatasan Republik Indonesia Tahun 2021: Penerapan Analisis Kluster K-Means dan Hierarki,” *JUSTIN (Jurnal Sist. dan Teknol. Informasi)*, vol. 12, no. 1, pp. 55–64, 2024, doi: 10.26418/justin.v12i1.69040.
15. A. Atira and B. Nurina Sari, “Penerapan Silhouette Coefficient, Elbow Method dan Gap Statistics untuk



- Penentuan Cluster Optimum dalam Pengelompokkan Provinsi di Indonesia Berdasarkan Indeks Kebahagiaan,” *J. Ilm. Wahana Pendidik.*, vol. 9, no. 17, pp. 76–86, 2023, [Online]. Available: <https://doi.org/10.5281/zenodo.8282638>.
16. A. F. Khairati, A. . Adlina, G. . Hertono, and B. . Handari, “Kajian Indeks Validitas pada Algoritma K-Means Enhanced dan K-Means MMCA,” *Prism. Pros. Semin. Nas. Mat.*, vol. 2, pp. 161–170, 2019, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/article/view/28906>.
  17. S. Pramana, R. Yordani, R. Kurniawan, and B. Yuniarto, *Dasar-dasar statistika dengan software R : konsep dan aplikasi*. Bogor: IN MEDIA, 2016.
  18. BPS, “Statistik Kriminal,” *Badan Pus. Stat.*, no. 021, pp. 1–62, 2023, [Online]. Available: <https://www.bps.go.id/id/publication/2023/12/12/5edba2b0fe5429a0f232c736/statistik-kriminal-2023.html>.