ISSN: 2339-2541
JURNAL GAUSSIAN, Volume 15, Nomor 1, Tahun 2026, Halaman 234 - 242
Online di: https://ejournal3.undip.ac.id/index.php/gaussian/

PENGELOMPOKAN DAERAH RAWAN KRIMINALITAS DI INDONESIA TAHUN 2021 MENGGUNAKAN METODE K-MEANS & FUZZY C-MEANS

Wulan Cahya Rahmadani[1*], Tatik Widiharih[2], Tarno[3]
[1,2,3] Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro
***e-mail** : wulancahyarahma@gmail.com

DOI: 10.14710/j.gauss.15.1.234-242

Article Info:
Received: 2024-08-25
Accepted: 2025-12-30
Available Online: 2026-08-12

Keywords:
Clustering ; k-means; fuzzy c-means; Davies Bouldin Index; Calinski Harabasz Index; Crime

Abstract: Crime is one of the aspects that can influence national stability and security. In this study, crime data is used to cluster crime-prone areas and is considered whether these areas require extra surveillance or not. This research employs the K-means and Fuzzy C-Means methods. The K-means method groups data based on the similarity of data with cluster centroids. This algorithm is relatively efficient in complexity and easy to understand. K-means explicitly allocates data to specific clusters. On the other hand, Fuzzy C-Means is capable of placing a data point that lies between two or more other clusters into a single cluster. This is due to each data point having a degree of membership to determine its grouping, making the chances of failure to converge or stable cluster centers very low. The optimal number of clusters is selected using validation through the Davies Bouldin Index and Calinski Harabasz Index. The research results indicate that clustering crime-prone areas using both methods yield the same outcome of 2 clusters. Cluster 1 exhibits a higher crime rate compared to cluster 2, as indicated by the higher average values of group members. The lowest Davies Bouldin Index is 1.04948, and the highest Calinski Harabasz Index is around 24.36783.

1. PENDAHULUAN

Suatu kehidupan ternyata tidak dapat terhindar dari perbuatan kejahatan atau kriminalitas. Kriminalitas merupakan perbuatan yang bertentangan dengan hukum, peraturan, serta norma yang berlaku di masyarakat. Kriminalitas merupakan salah satu faktor yang memiliki potensi untuk memengaruhi stabilitas dan keamanan nasional. Banyaknya kriminalitas yang terjadi di Indonesia, perlu dilakukan pengelompokan daerah rawan kriminalitas berdasarkan provinsi. Pengelompokan tersebut sebagai salah satu usaha untuk memberikan informasi kepada aparat penegak hukum dan dapat sebagai bahan pertimbangan dalam memutuskan apakah daerah tersebut memerlukan pengawasan ekstra atau tidak. Mayoritas tindak kriminal yang sering terjadi di Indonesia meliputi pelanggaran terhadap nyawa (pembunuhan), tindakan kekerasan fisik (penganiayaan), perbuatan melanggar moral (pemerkosaan/pencabulan), pencurian, peredaran narkoba, penipuan, dan juga tindak korupsi (Badan Pusat Statistik, 2022).

Pengelompokan adalah teknik pengolahan data statistik yang membantu mengelompokkan suatu wilayah menjadi *cluster* berdasarkan karakteristik tertentu. Pengelompokan daerah rawan kriminalitas ini menggunakan metode *k-means* dan *fuzzy c-means*. Metode *K-Means* mengelompokan data berdasarkan *centroid cluster* (MacQueen, 1967). Konsep dasar algoritma *K-means* yaitu meminimalkan *Sum of Square Error* (SSE) antara objek-objek data dengan sejumlah *k* centroid (Suyanto, 2019). Algoritma ini memiliki kompleksitas yang efisien dan mudah dipahami. *K-Means* secara tegas mengelompokkan data ke *cluster* tertentu. Sedangkan *Fuzzy C-Means* mampu menempatkan suatu data yang terletak diantara dua atau lebih *cluster* yang lain pada suatu *cluster* karena memiliki derajat keanggotaan (Kusumadewi dan Purnomo, 2010).

Langkah selanjutnya setelah *cluster* terbentuk adalah dengan menentukan jumlah *custer* optimum menggunakan validasi. Terdapat beberapa jenis validasi, salah satunya validasi relatif. Validasi relative digunakan untuk membandingkan berbagai metode pengelompokan atau untuk menentukan jumlah *cluster* yang optimal dalam suatu model. Baarsch dan Celebi (2012) menyatakan bahwa indeks validitas yang termasuk dalam kriteria relatif adalah indeks *Dunn, Silhouette, Davies-Bouldin , dan Calinski Harabasz* . Pada penelitian ini menggunakan validitas Davies Bouldin dan Calinski Harabasz. Jumlah *cluster* optimal dipilih berdasarkan nilai Indeks *Davies Bouldin* yang terendah (Davies dan Bouldin, 1979) serta nilai Indeks *Calinski Harabasz* yang tertinggi (Calinski dan Harabasz, 1974). Setelah sejumlah *cluster* optimum telah terbentuk, langkah selanjutnya adalah dengan melakukan profilisasi *cluster*. Profilisasi *cluster* digunakan untuk melihat nilai rata-rata dari setiap anggota masing-masing variabel di setiap *cluster*, sehingga didapatkan karakteristik setiap *cluster*.

Penelitian ini membahas tentang "Pengelompokan Daerah Rawan Kriminalitas di Indonesia tahun 2021 Menggunakan Metode *K-Means* dan *Fuzzy C-Means*". Penelitian ini bertujuan untuk mengetahui jumlah *cluster* optimal pada data kriminalitas dan menghasilkan *cluster* daerah rawan kriminalitas sebagai informasi bagi pemerintah untuk mempertimbangkan apakah daerah tersebut memerlukan pengawasan ekstra atau tidak.

2. TINJAUAN PUSTAKA

Analisis *cluster* menggunakan ukuran jarak akan bermasalah apabila variabel yang digunakan memiliki skala yang berbeda-beda. Cara yang paling sering digunakan untuk standarisasi variabel adalah menggunakan *Z-Score.* Proses ini mengubah data menjadi nilai standar, dengan rata-rata sebesar 0 dan standar deviasi sebesar 1 (Hair *et al.*, 2006).

Patro dan sahu (2015) menyatakan bahwa formula *Z-Score* sebagai berikut :

$$Z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j}$$

dengan:

Z_{ij} : objek ke i variabel ke j yang terlah distandarkan;

x_{ij} : objek ke i variabel ke j ;

Analisis *cluster* merupakan metode yang tepat untuk mengidentifikasi objek-objek yang homogen dan menggabungkannya ke dalam kelompok yang disebut *cluster*. Tujuan dari analisis *cluster* untuk mendapatkan *cluster* yang memiliki karakteristik yang sama. Analisis *cluster* tidak terlalu memperhitungkan asumsi dasar statistic tetap memperhatikan asumsi sampel representative dan multikolinieritas (Hair *et al.*, 2006)

a. Sampel Representatif

Asumsi ini digunakan untuk menentukan apakah sampel yang digunakan untuk analisis cukup atau cocok untuk analisis lebih lanjut. Asumsi ini dapat dihitung dengan menggunakan nilai KMO (Kaiser-Meyer-Olkin). Data dianggap memenuhi persyaratan kecukupan sampel jika nilai Indeks Kecekapan Sampel Kaiser-Meyer-Olkin (KMO) lebih besar dari 0,5 (Widarjono, 2015). Rumus untuk menghitung KMO adalah :

$$KMO = \frac{\sum_{j=1}^{p}\sum_{l=1\, j\neq l}^{p} r_{jl}^2}{\sum_{j=1}^{p}\sum_{l=1\, j\neq l}^{p} r_{jl}^2 + \sum_{i=1}^{p}\sum_{l=1\, j\neq l}^{p} a_{jl}^2} \quad (1)$$

dengan :

KMO : nilai Kaiser Mayer Olkin ; p : banyaknya variabel; n : banyaknya objek; x_{ij}: nilai objek ke-i pada variabel ke-j ; x_{ij} : nilai objek ke-i pada variabel ke-l ; r_{jl} : koefisien korelasi pearson antara variabel j dan l; $a_{ij,m}$: koefisien korelasi parsial antara variabel j dan l dengan menjaga agar variabel m konstan

b. Multikolinieritas

Tujuan dari uji multikolinearitas adalah untuk menguji apakah terdapat korelasi yang signifikan antara variabel bebas (independen) dalam suatu model regresi. Analisis *cluster* sebaiknya tidak terjadi multikolinearitas antar variabel. Salah satu cara untuk mendeteksi multikolinearitas adalah dengan menghitung Variance Inflation Factor (Ghozali, 2016). Rumus untuk menghitung Variance Inflation Factor (VIF) adalah :

$$\boldsymbol{VIF_j = \frac{1}{1- R_j^2}} \quad (2)$$

dengan VIF : nilai VIF untuk variabel ke j, j=1,2,…p ; R_j^2 : nilai koefisien determinasi variabel dependen (x_j) dengan variabel $x_k, k \neq j \; \forall k$.

Suatu data terindikasi multikolinieritas Jika nilai VIF ≥ 10, tindakan yang dapat diambil saat menghadapi multikolinearitas adalah dengan mengeluarkan variabel yang memiliki korelasi dalam model dan juga bisa menggunakan metode Principal Component Analysis (PCA).

K-Means Clustering

K-means cluster merupakan metode analisis klaster non-hierarkis yang berupaya membagi objek ke dalam satu atau lebih *cluster* berdasarkan karakteristik yang dimilikinya. Pada dasarnya, algoritma *K-Means* berfokus meminimalkan *Sum of Square Error* (SSE) antara objek-objek data dengan sejumlah k centroid (Suyanto, 2019). Algoritma *K-Means Cluster* ialah sebagai berikut :

1. Tentukan jumlah *cluster* k
2. Tentukan nilai *centroid cluster* awal sebanyak k secara random.
3. Proses iterasi untuk mencari anggota tiap *cluster* pada metode *k-means*

- Hitung jarak antara setiap objek dan *centroid* dari setiap *cluster* dengan menggunakan metode jarak Euclidean.
- Kelompokkan setiap objek ke *cluster* yang terdekat berdasarkan jarak objek tersebut dari pusat *cluster*.
- Hitunglah nilai *centroid* baru dengan cara mengambil rata-rata setiap objek dalam *cluster* yang bersangkutan.

$$\boldsymbol{v_{kj} = \frac{\sum_{i=1}^{n_k} x_{ij}}{n_k}} \quad (3)$$

dengan :

v_{kj}: *centroid* pada *cluster* ke-k variabel ke-j; $x_{i:}$ objek ke-i ; n_k: jumlah objek yang menjadi anggota *cluster* ke- k

- Proses iterasi dilakukan secara berulang hingga tidak ada perubahan pada anggota cluster.

Menurut Bezdek (1981), *fuzzy c-means* adalah metode *clustering* dimana setiap data mengikuti aturan himpunan *fuzzy.* Algoritma *fuzzy c-means* memiliki konsep dasar yaitu meminimumkan nilai fungsi objektif. Anggota *cluster* ditentukan oleh nilai derajat keanggotaan . Langkah-langkah *fuzzy c-means* sebagai berikut :

1. Menentukan data yang akan dikelompokkan dalam bentuk matriks dengan dimensi $n \times p$.
2. Menentukan beberapa input yang dibutuhkan:
- Jumlah *cluster* yang akan dibentuk (k>1)
- Pangkat pembobot (w>1)
- Maksimum iterasi (Maxiter)
- Error terkecil ($\varepsilon > 0$)
- Fungsi Objektif awal (P^0=0)
3. Bangkitkan bilangan acak ${U_{ik}}^0$, i =1,2,…, n ; k = 1,2,…,c sebagai elemen-elemen matriks U dengan syarat $\sum_{k=1}^{c} u_{ik} = 1$, $0 < u_{ik} < 1$ (Khang *et al.*, 2020)
4. Menghitung pusat *cluster* pada iterasi 1 : v_{kj}^1

$$\boldsymbol{v_{kj}^1} = \frac{\sum_{i=1}^{n}({U_{ik}}^0)^w x_{ij}}{\sum_{i=1}^{n}({U_{ik}}^0)^w} \quad (4)$$

dengan :
n : banyaknya data (i =1,2,…,n); w: bobot pangkat; v_{kj}^1: pusat *cluster* / centroid pada *cluster* ke-k variabel ke-j iterasi 1; ${U_{ik}}^0$: nilai derajat keanggotaan awal; x_{ij} : nilai objek ke-i variabel ke-j

5. Hitung nilai fungsi objektif pada iterasi 1 (P^1)

$$\boldsymbol{P^1} = \sum_{i=1}^{n}\sum_{k=1}^{c}\left({U_{ik}}^0\right)^w d_{ik}^2(x_i, v_k) \quad (5)$$

dengan :
P^1 : fungsi objektif pada iterasi ke-1; n: banyaknya data yang akan di *cluster* (i =1,2,…, n); c: jumlah *cluster* yang ditentukan; w: pangkat pembobot; u_0: nilai derajat keanggotaan awal; d_{ik}: jarak antara objek ke-i dan pusat *cluster* ke-k; x_i : nilai objek ke-i; v_k: pusat *cluster* ke-k

6. Menghitung ulang matriks keanggotaan u_{ik}

$$\boldsymbol{u_{ik}^1} = \left[\frac{\left[\sum_{j=1}^{p}(x_{ij}-v_{kj}^1)^2\right]^{\frac{1}{w-1}}}{\sum_{k=1}^{c}\left[\sum_{j=1}^{p}(x_{ij}-v_{kj}^1)^2\right]^{\frac{1}{w-1}}}\right]^{-1} \quad (6)$$

dengan :
p : variabel/atribut (j =1,2,..., p); c : banyaknya *cluster* (k =1,2,…, c); w : bobot pangkat; x_{ij} : objek ke-i dan variabel ke-j; v_{kj}^1 : pusat *cluster* ke-k dan variabel ke-j iterasi 1

7. Menghitung pusat *cluster* pada iterasi ke-t : v_{kj}

$$\boldsymbol{v_{kj}^t} = \frac{\sum_{i=1}^{n}(u_{ik})^w x_{ij}}{\sum_{i=1}^{n}(u_{ik})^w} \quad (7)$$

dengan :
n : banyaknya data (i =1,2,…,n); w : bobot pangkat; v_{kj}: pusat *cluster* / centroid pada *cluster* ke-k variabel ke-j ; u_{ik}: nilai derajat keanggotaan objek ke-i pada *cluster* ke-k; x_{ij} : nilai objek ke-i variabel ke-j

8. Menghitung nilai fungsi objektif pada iterasi ke-t (P^t)

$$\boldsymbol{P^t} = \sum_{i=1}^{n}\sum_{k=1}^{c}(u_{ik})^w d_{ik}^2(x_i, v_k) \quad (8)$$

dengan :

P^t : fungsi objektif pada iterasi ke-t; n : banyaknya data yang akan di *cluster* (i =1,2,…, n); c: jumlah *cluster* yang ditentukan; w : pangkat pembobot untuk derajat ke-fuzzy-an dengan w ϵ (1,∞); u_{ik}: nilai derajat keanggotaan objek ke-i pada *cluster* ke-k; d_{ik}: jarak antara objek ke-i dan pusat *cluster* ke-k; x_i : nilai objek ke-i; v_k: pusat *cluster* ke-k

Hitung ulang matriks keanggotaan u_{ik}

$$\boldsymbol{u_{ik}^t} = \left[\frac{\left[\sum_{j=1}^{p}(x_{ij}-V_{kj})^2\right]^{\frac{1}{w-1}}}{\sum_{k=1}^{c}\left[\sum_{j=1}^{p}(x_{ij}-V_{kj})^2\right]^{\frac{1}{w-1}}} \right]^{-1} \quad \textbf{(9)}$$

dengan :

p : variabel/atribut (j=1,2,...,p); c : banyaknya *cluster* (k=1,2,…,c); w : bobot pangkat; x_{ij} : objek ke-i dan variabel ke-j; V_{kj} : pusat *cluster* ke-k dan variabel ke-j

9. Memeriksa kondisi berhenti (konvergen)
   - Iterasi berhenti apabila $\left|P^{(t)} - P^{(t-1)}\right| < \varepsilon$ atau t > MaxIter
   - Jika tidak, maka t=t+1, mengulang kembali ke langkah 7
10. Jika iterasi berhenti, maka diperoleh hasil $u_{ik}\ final$. Penentuan anggota *cluster* dipilih berdasarkan nilai yang mendekati 1 atau tertinggi dari masing-masing baris dan masuk ke dalam *cluster* sesuai kolom.

Saat melakukan analisis *cluster* sering kali mengalami kesulitan dalam menentukan *cluster* optimum sehingga diperlukkan adanya teknik evaluasi *cluster*.

a. Davies Bouldin Index (DBI)

Pertama kali diperkenalkan oleh David L. Davies dan Donald W. Bouldin pada tahun 1979. Suatu *cluster* akan dianggap memiliki skema *cluster*ing yang optimal adalah yang memiliki nilai DBI minimal.

Langkah-langkah perhitungan Davies Bouldin Index adalah sebagai berikut :

1. Sum Of Square Within-*Cluster*(SSW)

Untuk mengetahui kohesi dalam sebuah *cluster* ke-i salah satunya adalah dengan menghitung nilai dari Sum Of Square Within-*Cluster* (SSW).

$$\boldsymbol{SSW_k} = \frac{\mathbf{1}}{\boldsymbol{n_k}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{n_k}} \boldsymbol{d(X_i, V_k)} \quad \textbf{(10)}$$

2. Sum Of Square Between-*Cluster* (SSB)

Tujuan dari menghitung Sum Of Square Between-*Cluster* (SSB) adalah memahami sejauh mana *cluster* antar kelompok berbeda satu sama lain atau seberapa besar jarak antar kelompok.

$$\boldsymbol{SSB_{ks}} = \boldsymbol{d(V_k, V_s), k \neq s} \quad (11)$$

dengan :

$d(v_k, v_s)$: jarak antara *centroid* ke-k dan *centroid* ke-s

3. Ratio (Rasio)

Perhitungan rasio ($R_{k,s}$) dilakukan dengan tujuan untuk memperoleh nilai perbandingan antara *cluster* ke-k dan *cluster* ke-s, dengan maksud mengukur nilai rasio setiap *cluster*.

$$\boldsymbol{R_{k,s}} = \frac{\mathbf{SSW_k + SSW_s}}{\mathbf{SSB_{k,s}}} \quad (12)$$

4. Davies Bouldin Index (DBI)

Nilai DBI diperoleh setelah mendapatkan nilai rasio dari persamaan (12). Semakin kecil nilai DBI, semakin tinggi kualitas pengelompokan *cluster* (Prasetyo, 2014).

$$\boldsymbol{DBI} = \frac{\mathbf{1}}{\boldsymbol{k}} \sum_{\boldsymbol{s=1}}^{\boldsymbol{k}} \boldsymbol{max_{k \neq s}(R_{k,s})} \quad \textbf{(13)}$$

b. Calinski Harabasz Index (CH Index)

Semakin besar nilai CH index maka semakin baik hasil *cluster* tersebut (Baarsch dan Celebi, 2012). Calinski Harabasz Index dapat dihitung sebagai berikut :

1. Menghitung nilai *centroid* global yang merupakan mean dari tiap variabel data

$$\bar{\boldsymbol{x}}_{\boldsymbol{j}} = \frac{\mathbf{1}}{\boldsymbol{n}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{x_{ij}} \quad (14)$$

dengan :

$\bar{x}_j$: rata-rata variabel ke j; x_{ij}: data ke-i; n : banyaknya data

2. Menghitung nilai Sum of Square Beetween-*cluster* (SSB)

$$\text{SSB} = \sum_{\boldsymbol{k=1}}^{\boldsymbol{K}} \boldsymbol{n_k d^2 (V_k, \bar{x})} \quad (15)$$

dengan :

$d^2(V_k, \bar{x})$: jarak kuadrat dari *cluster* ke-k dengan *centroid* global; n_k: banyaknya data pada *cluster* ke-k; V_k: *centroid* ke-k; $\bar{x}$: *centroid* global

3. Menghitung nilai Sum of Square Within-*cluster* (SSW)

$$\boldsymbol{SSW} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{k}} \sum_{\boldsymbol{x \in c_i}} \boldsymbol{d^2 (V_k, x_i)} \quad (16)$$

dengan :

$d^2(V_k, x)$: jarak kuadat dari *centroid* ke-k dengan setiap data; V_k : *centroid* ke-k; x_i : data ke-i

4. Menghitung CH Index dengan persamaan

$$\boldsymbol{CH} = \frac{\boldsymbol{n-k}}{\boldsymbol{k-1}} \times \frac{\boldsymbol{SSB}}{\boldsymbol{SSW}} \quad (17)$$

dengan :

n: Jumlah data ; k: Jumlah *cluster; SSB*: Sum of Square Beetween-*cluster*; *SSW*: Sum of Square Within-*cluster*

Tahap terakhir dari penelitian ini yaitu Profilisasi *cluster* digunakan untuk melihat nilai rata-rata dari setiap anggota masing-masing *cluster* yang akan menghasilkan karakteristik dari setiap *cluster*.

3. METODE PENELITIAN

Jenis data yang digunakan pada penelitian ini merupakan data sekunder. Data tersebut diperoleh dari buku profil Statistik Kriminalitas 2022 hasil publikasi Badan Pusat Statistik tahun 2022 (Badan Pusat Statistik, 2022) berupa data jenis kejahatan di Indonesia berdasarkan kriteria yang telah ditentukkan. Variabel yang digunakan yaitu Kejahatan terhadap nyawa (X1), Kejahatan terhadap fisik (X2), Kejahatan terhadap kesusilaan (X3), Kejahatan terhadap kemerdekaan (X4), Kejahatan terhadap hak milik (X5), Kejahatan terkait narkotika (X6), Kejahatan terkait penipuan, penggelapan, dan korupsi (X7). Tahapan analisis data sebagai berikut :

1. Input data berukuran $n \times 7$
2. Uji asumsi analisis *cluster*
3. Analisis data menggunakan *k-means clustering*
4. Analisis data menggunakan *fuzzy c-means clustering*
5. Evaluasi *cluster* menggunakan validasi *davies bouldin index* dan *calinski harabasz index*
6. Profilisasi *cluster*

4. HASIL DAN PEMBAHASAN

Data yang telah distandarkan menggunakan standarisasi *Z-Score,* kemudian diuji asumsi klaster, sebagai berikut :

a. Representatif

Uji asumsi representative tidak dilakukan karena melibatkan keseluruhan provinsi yang ada di Indonesia.

b. Multikolinieritas

Tabel 1 menunjukkan nilai VIF untuk seluruh variabel.

Tabel 1. Nilai VIF dan Setiap Variabel

Variabel	Nilai VIF
X_1	1,470
X_2	3,461
X_3	5,980
X_4	4,149
X_5	2,274
X_6	9,609
X_7	7,285

Tabel 1 memperlihatkan nilai VIF seluruh variabel di bawah 10. Kesimpulan yang dapat diperoleh berdasarkan Tabel 1 ialah tidak terdapat multikolinearitas antar variabel, sehingga dapat dilakukan analisis *cluster* lebih lanjut.

Hasil pengelompokan k=2,3…6; menggunakan metode *k-means cluster* dengan validasi *davies bouldin index* dan *calinski harabasz index* maka diperoleh nilai yang dapat dilihat pada Tabel2.

Tabel 2. Hasil Validasi Algoritma *K-Means*

Cluster	Validasi	
	Davies Bouldin Index	Calinski Harabasz Index
2	**1,04948**	**24,36783**
3	1,089275	18,79743
4	1,060656	14,76381
5	1,758399	12,59493
6	1,187008	13,2309

Berdasarkan Tabel 2, dengan metode *K-Means* diperoleh hasil bahwa pada pengelompokan daerah rawan kriminalitas nilai Davies Bouldin Index terendah yaitu 1,04948 dan nilai Calinski Harabasz Index tertinggi yaitu 24,36783 yang berada pada pengelompokan 2 *cluster.* Berdasarkan hasil peng*cluster*an k=2,3…6; menggunakan metode *fuzzy c-means cluster* dengan validasi *davies bouldin index* dan *calinski harabasz index* maka diperoleh nilai *davies bouldin index* dan *calinski harabasz index* pada Tabel 3.

Berdasarkan Tabel 3, diperoleh hasil bahwa pada pengelompokan daerah rawan kriminalitas 34 provinsi di Indonesia dengan metode *fuzzy c-means* diperoleh nilai *Davies Bouldin Index* terendah yaitu 1,04948 dan nilai *Calinski Harabasz Index* tertinggi yaitu 24,36783 yang berada pada pengelompokan 2 *cluster.*

Berdasarkan validasi *cluster* diperoleh bahwa jumlah *cluster* terbaik menurut K-Means *Cluster*ing dan Fuzzy C-Means dengan validasi Davies Bouldin Index dan Calinski Harabasz adalah 2 *cluster*. Karakteristik dari kelompok yang telah terbentuk dapat dijelaskan dengan cara melihat nilai rata-rata dari setiap variabel dalam penelitian. Tabel 4 menunjukkan nilai rata-rata setiap *cluster.*

Tabel 3. Perhitungan Nilai Validasi

Jumlah *Cluster*	Metode	Validasi	
		Davies Bouldin Index	Calinski Harabasz Index
2	K-Means	**1,04948**	**24,36783**
	Fuzzy C-Means	**1,04948**	**24,36783**
3	K-Means	1,089275	18,79743
	Fuzzy C-Means	1,376369	17,15107
4	K-Means	1,060656	14,76381
	Fuzzy C-Means	1,164955	19,3671
5	K-Means	1,758399	12,59493
	Fuzzy C-Means	1,217191	17,08857
6	K-Means	1,187008	13,2309
	Fuzzy C-Means	1,104375	15,9666

Tabel 4. Nilai rata-rata setiap *cluster*

Variabel	***Cluster* 1**	***Cluster* 2**
Kejahatan Terhadap Nyawa	**67,80**	20,27
Kejahatan Terhadap Fisik	**2.394,2**	555,86
Kejahatan Terhadap Kesusilaan	**356,2**	142,21
Kejahatan Terhadap Kemerdekaan	**137,8**	33,34
Kejahatan Terhadap Hak Milik	**6.554,6**	1.428,0
Kejahatan Terhadap Narkotika	**3.817,2**	616,83
Kejahatan Terhadap Penipuan, Penggelapan, dan Korupsi	**3.565,8**	595,31

Berdasarkan Tabel 4, diperoleh bahwa :

1. Seluruh nilai rata-rata tertinggi terdapat pada *cluster* 1, hal ini menandakan bahwa beberapa kejahatan di Indonesia sebagian besar terjadi pada provinsi yang terletak pada *cluster* 1 meliputi provinsi Sumatera Utara, DKI Jakarta, Sulawesi Selatan, Jawa Timur, dan Sumatera Selatan sehingga dikategorikan sebagai kelompok dengan daerah rawan kriminalitas tinggi. Provinsi yang menjadi anggota *cluster* 1 sebaiknya meningkatkan pengawasan dan keamanan di daerah tersebut.
2. *Cluster* 2 beranggotakan 29 provinsi yaitu Riau, Sulawesi Tengah, Jambi, Lampung, Kep. Riau, Jawa Tengah, Maluku Utara, Jawa Barat, Maluku, DI Yogyakarta, Bali, Nusa Tenggara Timur, Kalimantan Barat, Papua Barat, Kalimantan Timur, Bengkulu, Kalimantan Tengah, Kalimantan Selatan, Banten, Kalimantan Utara, Sulawesi Utara, Kep. Bangka Belitung, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Nusa Tenggara Barat, dan Papua dikategorikan sebagai kelompok dengan daerah rawan kriminalitas rendah.

5. KESIMPULAN

Metode *K-Means* dan *Fuzzy C-Means* memberikan hasil yang serupa dalam pengelompokan daerah rawan kriminalitas di Indonesia berdasarkan jenis kejahatan. Hasil dari penerapan validasi *cluster Davies Bouldin Index* dan *Calinski Harabasz Index* diperoleh nilai validasi yang sama yaitu jumlah *cluster* optimum sebanyak 2 *cluster* dengan nilai DBI terkecil sebesar 1,04948 dan nilai Calinski Harabasz Index tertinggi sebesar 24,36783. Karakteristik yang terbentuk dari masing-masing *cluster* berdasarkan algoritma *k-means* dan *fuzzy c-means* 2 *cluster* yaitu, *cluster* 1 menunjukkan nilai rata-rata jenis kejahatan yang lebih tinggi dibandingkan dengan *cluster* 2. Oleh karena itu, disarankan agar pemerintah mengintensifkan pengawasan dan keamanan di daerah yang termasuk dalam *cluster* 1.

DAFTAR PUSTAKA

Baarsch, J. dan Celebi, M.. (2012) 'Investigation of Internal Validity Measures for K-Means Clustering'. Available at: https://www.semanticscholar.org/paper/Investigation-of-Internal-Validity-Measures-for-Baarsch-Celebi/6dd2a2d3fb9dddf028283f060fb79a6c1edc9c61.

Badan Pusat Statistik (2022) *Statistik Kriminal 2022*. Available at: https://www.bps.go.id/publication/2022/11/30/4022d3351bf3a05aa6198065/statistik-kriminal-2022.html.

Bezdek, J.C. (1981) 'Pattern Recognition With Fuzzy Objective Function Algorithms'.

Calinski, T. dan Harabasz, J. (1974) 'A dendrite method for cluster analysis', *Communications in Statistics - Theory and Methods*, 3(1), pp. 1–27. Available at: https://doi.org/10.1080/03610927408827101.

Davies, D.L. dan Bouldin, D.W. (1979) 'AClusterSeparationMeasure', pp. 224–227.

Hair, J.F., Black, W.C., Babbin, J.B. dan Anderson, R.E. (2006) *Multivariate Data Analysis (7th Edition)*. Available at: https://www.pdfdrive.com/multivariate-data-analysis-7th-edition-d156708931.html.

Khang, T.D., Vuong, N.D., Tran, M.-K. dan Fowler, M. (2020) 'Fuzzy C-Means Clustering Algorithm with Multiple Fuzzification Coefficients', *Algorithms*, 13(7), p. 158. Available at: https://doi.org/10.3390/a13070158.

Kusumadewi, S. dan Purnomo, H. (2010) *Aplikasi Logika Fuzzy untuk pendukung keputusan*.

MacQueen, J.B. (1967) 'Some Methods For Classification and Analysis of Multivariate Observations'. Available at: https://www.semanticscholar.org/paper/Some-methods-for-classification-and-analysis-of-MacQueen/ac8ab51a86f1a9ae74dd0e4576d1a019f5e654ed.

Patro, S.G.K. dan sahu, K.K. (2015) 'Normalization: A Preprocessing Stage', *Iarjset*, 2(3), pp. 20–22. Available at: https://doi.org/10.17148/iarjset.2015.2305.

Prasetyo, E. (2014) *Data Mining Mengolah Data menjadi Informasi menggunakan Matlab*. Yogyakarta: ANDI.

Suyanto (2019) *Data Mining untuk Klasifikasi dan Klasterisasi Data*. Bandung: Informatika.