Jurnal Komputer dan Informatika
Vol 19 No 2, Oktober 2024: hlm 83-94

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

PENENTUAN ALGORITMA TERBAIK DALAM ANALISIS KLASTER TINDAK PIDANA DI INDONESIA MENGGUNAKAN *K-MEANS* DAN *FUZZY C-MEANS*

Derren Fusta
Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara,
Jln. Letjen S. Parman No. 1, Jakarta, 11440, Indonesia
E-mail: derren.535210094@stu.untar.ac.id

ABSTRAK

Tindak pidana merupakan salah satu isu yang berdampak signifikan terhadap keamanan dan stabilitas masyarakat. Banyak faktor yang mempengaruhi terjadinya tindak pidana, seperti kondisi sosial, ekonomi, dan lingkungan. Untuk mengidentifikasi tindak pidana yang sudah terjadi, teknologi machine learning dapat digunakan untuk mengidentifikasi pola dalam data tindak pidana. Algoritma *K-Means* dan *Fuzzy C-Means* merupakan algoritma *machine learning* yang digunakan dalam penelitian ini untuk mengidentifikasi pola dalam data tindak pidana menggunakan metode analisis klaster berdasarkan jumlah tindak pidana di suatu daerah, sehingga nantinya dapat mempermudah pemerintah dan pihak berwenang setempat dalam proses mengambil kebijakan atau tindakan lebih lanjut. Dataset yang digunakan dalam penelitian ini berasal dari kepolisian daerah negara Indonesia tahun 2000 hingga tahun 2022 yang ditemukan pada website Badan Pusat Statistik Indonesia. Penelitian ini bertujuan untuk menentukan algoritma terbaik terhadap hasil kinerja algoritma *K-Means* dan *Fuzzy C-Means* dalam analisis klaster tindak pidana di Indonesia. Hasil pengujian dengan menggunakan metode evaluasi Koefisien *Silhouette* dan *Davies-Bouldin Index* menunjukkan algoritma K-Means dan algoritma *Fuzzy C-Means* mendapat hasil nilai yang serupa yaitu sebesar 0.7078 pada Koefisien *Silhouette* dan sebesar 0.5457 pada *Davies-Bouldin Index*.

Kata kunci—*K-Means, Fuzzy C-Means, Machine Learning*, Tindak Pidana

ABSTRACT

Crime is an issue that has a significant impact on the security and stability of society. Many factors influence the occurrence of criminal acts, such as social, economic and environmental conditions. To identify criminal acts that have occurred, machine learning technology can be used to identify patterns in criminal act data. The K-Means and Fuzzy C-Means algorithms are machine learning algorithms used in this research to identify patterns in criminal act data using a cluster analysis method based on the number of criminal acts in an area, so that later it can make it easier for the local government and police in the process of making policies or further action. The dataset used in this research comes from the Indonesian regional police from 2000 to 2022 which was found on the website of the Indonesian Central Bureau of Statistics. This research aims to determine the best algorithm for the performance results of the K-Means and Fuzzy C-Means algorithms in cluster analysis of criminal acts in Indonesia. Test results using the Silhouette Coefficient and Davies-Bouldin Index evaluation methods show that the K-Means algorithm and the Fuzzy C-Means algorithm obtained similar results, namely 0.7078 for the Silhouette Coefficient and 0.5457 for the Davies-Bouldin Index.

Keywords*—*K-Means, Fuzzy C-Means, Machine Learning, Crime

1. PENDAHULUAN

Tindak pidana merupakan permasalahan yang memiliki dampak signifikan terhadap stabilitas sosial dan keamanan di suatu negara. Pada negara Indonesia, tindak pidana yang dilakukan oleh seseorang mempunyai tingkatan yang berbeda-beda setiap tahunnya [1]. Beberapa jenis tindak pidana, seperti pencurian, kekerasan, penipuan, hingga kejahatan siber dapat merugikan orang lain secara moral atau materil serta melanggar undang-undang dan norma-norma yang ada di suatu daerah [2]. Dampak yang dirasakan setelah melakukan tindak pidana itu, tidak hanya dirinya sendiri tetapi juga keluarganya merasakan dampak itu [3]. Latar belakang yang membuat seseorang melakukan

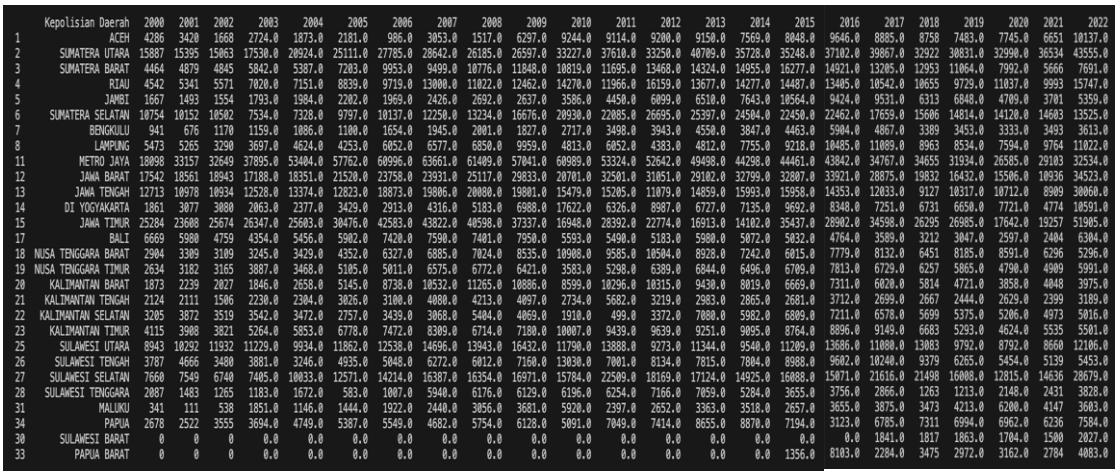
tindak pidana biasanya dipengaruhi oleh kondisi ekonomi, kondisi sosial, dan kondisi psikologis atau tekanan mental seseorang [4]. Dari sekian banyaknya tindak pidana yang dapat dilakukan oleh seseorang, tindak pidana yang sering dilakukan oleh seseorang saat ini ialah pembunuhan, baik pembunuhan terencana atau tidak [4]. Pelaku itu tidak berpikir panjang terhadap konsekuensi yang didapatkannya saat melakukan perbuatan itu, sehingga pelaku tidak segan untuk membunuh korban [3]. Padahal, perbuatan itu bisa menyebabkan trauma yang membekas pada keluarga korban atau orang-orang terdekat korban [5]. Tindak pidana yang dilakukan oleh pelaku biasanya dilakukan lebih dari 1 kali, jika pelaku belum diketahui oleh publik atau masuk ke sel tahanan [6]. Oleh karena itu, kepolisian perlu mengatur strategi atau pengawasan yang efektif pada daerah yang rawan melakukan tindak pidana, agar orang yang mau melakukan tindak pidana berpikir 2 kali sebelum bertindak dan dapat dicegah atau ditangani dengan cepat jika ada yang melakukan tindak pidana. Dengan analisis klaster pada data tindak pidana menggunakan algoritma K-Means dan Fuzzy C-Means, maka diharapkan dapat membantu kepolisian dalam menemukan pola dan mengelompokkan daerah yang rawan melakukan tindak pidana berdasarkan jumlah orang yang melakukan tindak pidana di suatu daerah pertahun.

Penelitian terdahulu yang dilakukan oleh Adjie Dwi Sulistyono pada tahun 2024, tentang "Analisis Cluster Menggunakan Algoritma K-Means dan Centroid Linkage pada Data Kriminalitas Indonesia Tahun 2021", dari penelitian tersebut didapat hasil dari klastering yang dilakukan memiliki hasil *Silhouette* terbaik kira-kira sebesar 65% dengan menggunakan nilai n_clusters=2 [2]. Sedangkan penelitian terdahulu yang dilakukan oleh Resti Noor Fahmi pada tahun 2021, tentang "Analisis Pemetaan Tingkat Kriminalitas di Kabupaten Karawang Menggunakan Algoritma K-means", dari penelitian tersebut didapat hasil *Silhouette* terbaik sebesar 0.52% pada tahun 2019 dan sebesar 0.54 pada tahun 2020 [7].

2. METODE PENELITIAN

Pada penelitian analisis klaster tindak pidana ini, *machine learning* digunakan untuk mengelompokkan daerah di Indonesia berdasarkan jumlah tindak pidana dari kepolisian daerah di 28 provinsi yang tersebar diseluruh Indonesia dengan catatan data polda Kepulauan Riau tahun 2000-2022 merupakan bagian dari polda Riau, data polda Kep. Bangka Belitung tahun 2000-2022 merupakan bagian dari polda Sumatera Selatan, data polda Banten tahun 2000-2022 merupakan bagian dari polda Jawa Barat, data polda Gorontalo tahun 2000-2022 merupakan bagian dari polda Sulut, data polda Maluku Utara tahun 2000-2022 merupakan bagian dari polda Maluku, data polda Metro Jaya meliputi polres Jakarta Selatan, Jakarta Timur, Jakarta Pusat, Jakarta Utara, Jakarta Barat, Kepulauan Seribu, Kabupaten Bekasi, Kota Bekasi, Kabupaten Tangerang, Kota Tangerang, Kota Depok, Bandara Soekarno-Hatta, dan Kesatuan Pelaksanaan Pengamanan Pelabuhan (KP3) Tanjung Priok, data polda Kalimantan Utara tahun 2000-2022 merupakan bagian dari Polda Kalimantan Timur. Ada 2 algoritma *machine learning* yang digunakan dalam proses analisis klasternya, yaitu: K-Means dan Fuzzy C-Means. Kedua algoritma ini digunakan tidak hanya untuk analisis klaster data tindak pidana, tetapi juga untuk menentukan algoritma mana yang terbaik berdasarkan hasil nilai evaluasinya dalam menganalisis klaster data tindak pidana. Dikarenakan total file yang diunduh pada website Badan Pusat Statistik Indonesia terpisah dan dibagi menjadi 8 bagian yang berisikan jumlah data tindak pidana di seluruh provinsi Indonesia masing-masing setiap 3 tahun sekali, mulai dari tahun 2000-2002, tahun 2003-2005, tahun 2006-2008, tahun 2009-2011, tahun 2012-2014, tahun 2015-2017, tahun 2018-2020, dan tahun 2021-2022, maka dilakukan proses *concat* untuk menggabungkan 8 bagian file yang diunduh menjadi 1 bagian file saja yang berisikan jumlah data tindak pidana di seluruh provinsi Indonesia mulai dari tahun 2000-2022. Hasil dari proses *concat* tersebut dapat dilihat pada Gambar (1).

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)



	Kepolisian Daerah	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
1	ACEH	4286	3420	1668	2724.0	1873.0	2181.0	986.0	3053.0	1517.0	6297.0	9244.0	9114.0	9200.0	9150.0	7569.0	8048.0	9646.0	8885.0	8758	7483.0	7745.0	6651	10137.0
2	SUMATERA UTARA	15887	15395	15063	17530.0	20924.0	25111.0	27785.0	28642.0	26185.0	26597.0	33227.0	37610.0	33250.0	40709.0	35728.0	35248.0	37102.0	39867.0	32922	30831.0	32990.0	36534	43555.0
3	SUMATERA BARAT	4464	4879	4845	5842.0	5387.0	7203.0	9953.0	9499.0	10776.0	11848.0	10819.0	11695.0	13468.0	14324.0	14955.0	16277.0	14921.0	13205.0	12953	11064.0	7992.0	5666	7691.0
4	RIAU	4542	5341	5571	7020.0	7151.0	8839.0	9719.0	13000.0	11022.0	12462.0	14270.0	11966.0	16159.0	13677.0	14277.0	14487.0	13405.0	10542.0	10655	9729.0	11037.0	9993	15747.0
5	JAMBI	1667	1493	1554	1793.0	1984.0	2202.0	1969.0	2426.0	2692.0	2637.0	3586.0	4450.0	6099.0	6510.0	7643.0	10564.0	9424.0	9531.0	6313	6848.0	4709.0	3701	5359.0
6	SUMATERA SELATAN	10754	10152	10502	7534.0	7328.0	9797.0	10137.0	12250.0	13234.0	16676.0	20930.0	22085.0	26695.0	25397.0	24504.0	22450.0	22462.0	17659.0	15606	14814.0	14120.0	14603	13525.0
7	BENGKULU	941	676	1170	1159.0	1086.0	1100.0	1654.0	1945.0	2001.0	1827.0	2717.0	3498.0	3943.0	4550.0	3847.0	4463.0	5904.0	4867.0	3389	3453.0	3333.0	3493	3613.0
8	LAMPUNG	5473	5265	3290	3697.0	4624.0	4253.0	6052.0	6577.0	6850.0	9959.0	4813.0	6052.0	4383.0	4812.0	7755.0	9218.0	10485.0	11089.0	8963	8534.0	7594.0	9764	11022.0
11	METRO JAYA	18098	33157	32649	37895.0	53404.0	57762.0	60996.0	63661.0	61409.0	57041.0	60989.0	53324.0	52642.0	49498.0	44298.0	44461.0	43842.0	34767.0	34655	31934.0	26585.0	29103	32534.0
12	JAWA BARAT	17542	18561	18943	17188.0	18351.0	21520.0	23758.0	23931.0	25117.0	29833.0	20701.0	32501.0	31051.0	29102.0	32799.0	32807.0	33921.0	28875.0	19832	16432.0	15506.0	10936	34523.0
13	JAWA TENGAH	12713	10978	10934	12528.0	13374.0	12823.0	18873.0	19806.0	20080.0	19801.0	15479.0	15205.0	11079.0	14859.0	15993.0	15958.0	14353.0	12033.0	9127	10317.0	10712.0	8909	30060.0
14	DI YOGYAKARTA	1861	3077	3080	2063.0	2377.0	3429.0	2913.0	4316.0	5183.0	6988.0	17622.0	6326.0	8987.0	6727.0	7135.0	9692.0	8348.0	7251.0	6731	6650.0	7721.0	4774	10591.0
15	JAWA TIMUR	25284	23608	25674	26347.0	25603.0	30476.0	42583.0	43822.0	40598.0	37337.0	16948.0	28392.0	22774.0	16913.0	14102.0	35437.0	28902.0	34598.0	26295	26985.0	17642.0	19257	51985.0
17	BALI	6669	5980	4759	4354.0	5456.0	5902.0	7420.0	7590.0	7401.0	7950.0	5593.0	5490.0	5183.0	5980.0	5072.0	5032.0	4764.0	3589.0	3212	3047.0	2597.0	2404	6304.0
18	NUSA TENGGARA BARAT	2904	3309	3109	3245.0	3429.0	4352.0	6327.0	6885.0	7024.0	8535.0	10900.0	9585.0	10504.0	8928.0	7242.0	6015.0	7779.0	8132.0	6451	8185.0	8591.0	6296	5296.0
19	NUSA TENGGARA TIMUR	2634	3182	3165	3887.0	3468.0	5105.0	5011.0	6575.0	6772.0	6421.0	3583.0	5298.0	6389.0	6844.0	6496.0	6709.0	7813.0	6729.0	6257	5865.0	4790.0	4909	5991.0
20	KALIMANTAN BARAT	1873	2239	2027	1846.0	2658.0	5145.0	8738.0	10532.0	11265.0	10886.0	8599.0	10296.0	10315.0	9430.0	8019.0	6669.0	7311.0	6020.0	5814	4721.0	3858.0	4048	3975.0
21	KALIMANTAN TENGAH	2124	2111	1506	2230.0	2304.0	3026.0	3100.0	4080.0	4213.0	4097.0	2734.0	5682.0	3219.0	2983.0	2865.0	2681.0	3712.0	2699.0	2667	2444.0	2629.0	2399	3189.0
22	KALIMANTAN SELATAN	3205	3872	3519	3542.0	3472.0	2757.0	3439.0	3068.0	5404.0	4069.0	1910.0	499.0	3372.0	7080.0	5982.0	6809.0	7211.0	6578.0	5699	5375.0	5206.0	4973	5016.0
23	KALIMANTAN TIMUR	4115	3908	3821	5264.0	5853.0	6778.0	7472.0	8309.0	6714.0	7180.0	10007.0	9439.0	9639.0	9251.0	9095.0	8764.0	8896.0	9149.0	6683	5293.0	4624.0	5535	5501.0
25	SULAWESI UTARA	8943	10292	11932	11229.0	9934.0	11862.0	12538.0	14696.0	13943.0	16432.0	11790.0	13888.0	9273.0	11344.0	9540.0	11209.0	13686.0	11080.0	13083	9792.0	8792.0	8660	12106.0
26	SULAWESI TENGAH	3787	4666	3480	3881.0	3246.0	4935.0	5048.0	6272.0	6012.0	7160.0	13030.0	7001.0	8134.0	7815.0	7804.0	8988.0	9602.0	10240.0	9379	6265.0	5454.0	5139	5453.0
27	SULAWESI SELATAN	7660	7549	6740	7405.0	10033.0	12571.0	14214.0	16387.0	16354.0	16971.0	15784.0	22509.0	18169.0	17124.0	14925.0	16088.0	15071.0	21616.0	21498	16008.0	12815.0	14636	28679.0
28	SULAWESI TENGGARA	2087	1483	1265	1183.0	1672.0	583.0	1007.0	5940.0	6176.0	6129.0	6196.0	6254.0	7166.0	7059.0	5284.0	3655.0	3756.0	2866.0	1263	1213.0	2148.0	2431	3828.0
31	MALUKU	341	111	538	1851.0	1146.0	1444.0	1922.0	2440.0	3056.0	3681.0	5920.0	2397.0	2652.0	3363.0	3518.0	2657.0	3655.0	3875.0	3473	4213.0	6200.0	4147	3683.0
34	PAPUA	2678	2522	3555	3694.0	4749.0	5387.0	5549.0	4682.0	5754.0	6128.0	5091.0	7049.0	7414.0	8655.0	8870.0	7194.0	3123.0	6785.0	7311	6994.0	6962.0	6236	7584.0
30	SULAWESI BARAT	0	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1841.0	1817	1863.0	1704.0	1500	2027.0
33	PAPUA BARAT	0	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1356.0	8103.0	2284.0	3475	2972.0	3162.0	2784	4083.0

Gambar 1 Jumlah Data Tindak Pidana Tahun 2000-2022 Di Indonesia

Dataset diperoleh melalui website Badan Pusat Statistik Indonesia. Setelah diunduh dari website tersebut, dataset tersebut terdapat 8 bagian data tindak pidana sejak tahun 2000-2022 yang dibagi menjadi masing-masing 3 tahun sekali setiap bagian data tindak pidana dan memiliki baris dan kolom yang bervariasi. Setelah dilakukan proses *concat*, maka 8 bagian data yang terpisah digabungkan menjadi 1 file utuh yang memiliki 28 baris dan 24 kolom. Pada algoritma K-Means dan Fuzzy C-Means, *n_clusters* diatur bervariasi agar dapat melihat dan menentukan hasil nilai *Silhouette* dan *Davies-Bouldin Index* terbaik saat *n_clusters* diatur ke angka tertentu.

2.1 K-Means

Sebelum dinamakan algoritma K-Means yang diperkenalkan oleh James MacQueen pada tahun 1967, konsep dasar algoritma itu diusulkan oleh Hugo Steinhaus pada tahun 1956 yang kemudian diajukan dalam bentuk standar oleh Stuart Lloyd dari *Bell Labs* pada tahun 1957 sebagai metode untuk modulasi kode pulsa [8][9][10]. K-Means merupakan algoritma *unsupervised learning* yang digunakan untuk klasterisasi data. Setiap klaster memiliki titik pusat (*centroid*) yang mewakili klaster tersebut [11]. Algoritma tersebut sering digunakan untuk klasterisasi dikarenakan algoritma itu paling sederhana dan cepat dalam mengelompokkan suatu data dengan jumlah yang besar [12]. Syarat untuk mengelompokkan suatu data pada algoritma itu yaitu memiliki kemiripan pada kriteria, kondisi, atau karakteristik antara satu data dengan data yang lain [13]. Oleh karena itu, prinsip dasar dari algoritma K-Means yaitu membagi sekumpulan data menjadi k bagian dengan menentukan pusat (*centroid*) atau rata-rata (*mean*) untuk setiap klaster [11]. Algoritma K-Means memiliki 3 parameter yang dapat dimodifikasi oleh pengguna untuk mendapatkan hasil klastering yang lebih baik yaitu jumlah cluster (*n_clusters*), inisialisasi klaster (*init*), dan jarak (*algorithm*) [12]. Untuk jumlah cluster (*n_clusters*) dapat diinput dengan angka yang bertipe data *integer*, inisialisasi klaster (*init*) dapat diinput dengan *k-means++* atau *random*, dan jarak (*algorithm*) dapat diisi dengan *lloyd*, *elkan*, *auto*, atau *full*. Berikut ini merupakan langkah-langkah tahapan algoritma K-Means dalam mengelompokkan suatu data adalah sebagai berikut [11][13][14]:

1. Tentukan jumlah klaster (*n_clusters*) pada dataset
2. Tentukan titik pusat (*centroid*) sebanyak *n_clusters*
3. Hitung jarak terdekat *centroid* menggunakan *Euclidean Distance*
4. Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan *centroid*
5. Hitung dan tentukan pusat klaster baru dengan anggota klaster yang sekarang dengan menggunakan persamaan (1), dimana:
   V_{ij}= *Centroid* rata-rata klaster ke-i untuk variabel ke-j
   N_i= Jumlah anggota klaster ke-i
   I,k= Indeks dari klaster
   J = Indeks dari variabel
   X_{kj}= Nilai data ke-k variabel ke-j untuk klaster tersebut
6. Terapkan kembali iterasi langkah ke-3 sampai ke-5 hingga *centroid* bernilai optimal

2.1.1 *Jarak Euclidean*

Jarak *Euclidean* adalah metode pengukuran jarak kartesian antara dua titik yang terletak dalam ruang hiper, yang bisa berupa bidang dua dimensi, tiga dimensi, atau ruang berdimensi lebih tinggi. Jarak itu dapat divisualisasikan sebagai panjang garis lurus yang menghubungkan kedua titik yang sedang dianalisis. Metrik itu sangat berguna untuk menghitung perubahan posisi atau perpindahan bersih antara dua kondisi objek yang dibandingkan [14]. Pada penelitian ini, jarak *Euclidean* digunakan untuk menghitung jarak terdekat *centroid*. Jarak *Euclidean* dihitung menggunakan persamaan (1), dimana:

d = jarak antara x dan y
x = data pusat klaster
y = data pada atribut
i = setiap data
n = jumlah data
xi = data pada pusat klaster ke-i
yi = data pada setiap data ke-i

$$d(x,y) = |x - y| = \sqrt{\Sigma_{i=1}^{N}(xi - yi)^2} \quad (1)$$

Jarak *Euclidean* memiliki kelebihan dan kekurangan dibandingkan dengan metode perhitungan jarak yang lain. Kelebihan pada jarak *Euclidean* terletak pada perhitungan jaraknya yang lebih umum digunakan pada algoritma K-Means dan hasil yang diperoleh lebih optimal dibandingkan dengan metode perhitungan jarak yang lain, sedangkan untuk kekurangan pada jarak *Euclidean* terletak pada waktu yang dibutuhkan untuk menghitung iterasi secara manual cenderung lebih lama disebabkan oleh jumlah iterasi pada jarak *Euclidean* yang lebih banyak dibandingkan dengan metode perhitungan jarak yang lain [14].

2.1.2 *K-Means++*

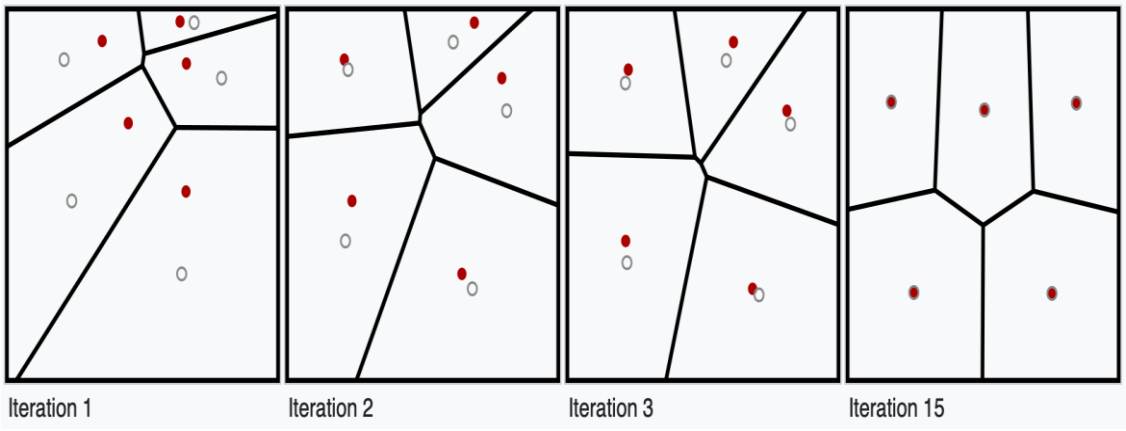
K-Means++ pertama kali diusulkan oleh David Arthur dan Sergei Vassilvitskii pada tahun 2007 sebagai algoritma perkiraan untuk menghindari pengelompokan yang terkadang buruk pada algoritma K-Means dan sebagai algoritma untuk memilih nilai awal atau *seed* untuk algoritma klaster K-Means. Algoritma K-Means++ menggunakan metode untuk menginisialisasi pusat klaster sebelum melanjutkan dengan proses optimasi menggunakan algoritma K-Means standar. Dengan menggunakan inisialisasi K-Means++, algoritma itu memastikan bahwa solusi yang ditemukan akan memiliki kinerja yang kompetitif [15][16].

2.1.3 *Lloyd*

Algoritma Lloyd pertama kali diusulkan oleh Stuart P. Lloyd dari Bell Labs pada tahun 1957 sebagai teknik modulasi kode pulsa yang digunakan untuk menemukan himpunan titik-titik yang berjarak sama dalam himpunan bagian ruang *Euclidean*. Sama seperti algoritma K-Means, algoritma ini secara berulang menghitung pusat massa dari setiap klaster dalam partisi, kemudian memperbarui pembagian data berdasarkan pusat massa yang paling dekat. Dalam pengaturan ini, operasi rata-rata berfungsi sebagai integral dalam ruang tertentu sementara penentuan pusat massa terdekat menghasilkan diagram Voronoi. Algoritma Lloyd dapat digunakan untuk berbagai keperluan, seperti kuantisasi, dithering, dan stippling [17][18].

Pada gambar 2 dibawah, contoh algoritma Lloyd dimulai pada iterasi 1 yang dimana titik-titik merah adalah pusat kluster (*centroid*) yang dipilih secara acak dan titik hitam merupakan data yang dikelompokkan ke centroid terdekat menggunakan batas-batas Voronoi diagram. Pada iterasi ke 2,

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)



Gambar 2 Contoh algoritma Lloyd

Centroid diperbarui berdasarkan rata-rata posisi titik-titik data yang termasuk dalam klusternya dan batas-batas kluster diperbarui sesuai jarak ke *centroid* baru. Pada iterasi ke 3, proses perbaruan *centroid* dan batas-batas kluster diulang dan titik data yang dikelompokkan berubah sedikit seiring pembaruan posisi centroid. Pada iterasi ke 15, algoritma berhenti ketika posisi centroid stabil dan tidak berubah lagi serta klaster telah terbentuk secara optimal.

2.1.4 Elkan

Algoritma Elkan pertama kali diusulkan oleh Charles Elkan pada tahun 2003 yang digunakan untuk meningkatkan efisiensi perhitungan jarak antara titik data dan centroid untuk mengoptimalkan algoritma K-Means. Algoritma Elkan merupakan variasi dari algoritma Lloyd yang memanfaatkan pertidaksamaan segitiga untuk mengurangi jumlah perhitungan jarak saat menetapkan titik ke klaster. Meskipun lebih cepat dari metode Lloyd, algoritma Elkan memerlukan penyimpanan yang sebanding dengan jumlah klaster dan titik data, sehingga kurang efisien jika jumlah klaster sangat besar [19]. Prinsip dari algoritma Elkan adalah, jika pembaruan pusat tidak mengubah posisinya secara signifikan, maka banyak perhitungan jarak antara titik dan pusat dapat dihindari saat penugasan titik ke pusat dihitung ulang. Untuk menentukan jarak mana yang perlu dihitung, pertidaksamaan segitiga digunakan untuk menetapkan batas bawah dan batas atas setelah pembaruan pusat [19].

2.2 Fuzzy C-Means

Fuzzy C-Means pertama kali dikembangkan oleh J.C. Dunn pada tahun 1973 yang kemudian dikembangkan lebih lanjut oleh J.C. Bezdek pada tahun 1981 [20][21]. Fuzzy C-Means merupakan algoritma *unsupervised learning* yang digunakan untuk klasterisasi data dimana keberadaan tiap titik-titik ditentukan oleh derajat keanggotaan untuk menentukan klaster yang optimal [22]. Derajat keanggotaan dikelompokkan dengan angka antara 0 dan 1 yang menunjukkan keanggotaan parsial dari data tersebut [23]. Prinsip dasar dari algoritma Fuzzy C-Means adalah dimulai dengan menentukan pusat klaster yang dimana menggambarkan posisi rata-rata masing-masing klaster dengan hasil awal pusat klaster belum akurat. Dengan proses iterasi yang berulang, dimana tingkat keanggotaan diperbarui untuk setiap titik data dan pusat klaster, maka posisi pusat klaster akan bergerak menuju lokasi yang lebih tepat. Proses iterasi ini bertujuan untuk meminimalkan fungsi tujuan yang mengukur jarak antara titik data dan pusat klaster dengan mempertimbangkan bobot yang diberikan oleh derajat keanggotaan titik data [24]. Keunggulan algoritma Fuzzy C-Means terletak pada kemampuannya dalam mengelompokkan data dengan dimensi yang beragam ke beberapa klaster yang berbeda dikarenakan algoritma itu mengizinkan keanggotaan bertingkat untuk tiap titik data, serta mampu menetapkan klaster yang bervariasi dan terukur [21][25]. Berikut ini merupakan langkah-langkah tahapan algoritma Fuzzy C-Means dalam mengelompokkan suatu data adalah sebagai berikut [22][24]:

1. Input data (X_{ij}) yang akan dikelompokkan ke dalam sebuah matriks berukuran $s \times p$, dengan s adalah banyaknya data yang akan dikelompokkan dan p adalah banyaknya atribut atau variabel.

2. Menentukan:
   a) Jumlah Klaster = c
   b) Pangkat = $w(>1)$
   c) Maksimal Iterasi = *maxIter*
   d) *Error T*erkecil = ξ
   e) Fungsi Objektif Awal = $P_o - 0$
   f) Iterasi Awal = 1
3. Membangkitkan bilangan random μ_{ik} antara 0 sampai 1 sebagai elemen matriks keanggotaan awal U.

$$U^{(0)} = \left[\mu_{11} \cdots \mu_{1c} \vdots \ddots \vdots \mu_{s1} \cdots \mu_{sc}\right] \quad (2)$$

   Hitung jumlah setiap baris:

$$Q_i = \sum_{k=1}{}^{c} \mu_{ik} \quad (3)$$

   Dengan i = 1,2, s Hitung:

$$\mu_{ik}{}^{(0)} = \frac{\mu_{ik}}{Q_i} \quad (4)$$

   Pada persamaan (2) dan (3), dimana:
   Q_i = Total nilai keanggotaan untuk data ke-i
   c = Jumlah klaster
   μ_{ik} = Nilai keanggotaan data ke-i terhadap klaster ke-k dengan nilai berada dalam rentang [0,1]
   k = Indeks klaster dari 1 hingga c (pada persamaan 2)
4. Hitung pusat klaster ke-k menggunakan persamaan (4), dimana:
   V_{kj} = Pusat klaster
   μ_{ik} = Derajat keanggotaan data ke-i terhadap klaster ke-k
   X_{ij} = Nilai data ke-i pada dimensi j
   w = Parameter *fuzziness* dengan $w > 1$

$$V_{kj} = \frac{\sum_{t=1}^{s} \left((\mu_{ik})^w X_{ij}\right)}{\sum_{t=1}^{s} (\mu_{ik})^w} \quad (5)$$

5. Memperbaiki derajat keanggotaan setiap data pada setiap klaster menggunakan persamaan (5), dimana:
   μ_{ik} = Derajat keanggotaan data ke-i terhadap klaster ke-k
   X_{ik} = Nilai data ke-i pada dimensi j
   V_{kj} = Pusat klaster ke-k pada dimensi j
   w = Parameter *fuzziness* dengan $w > 1$

$$\mu_{ik} = \frac{\left[\sum_{j=1}^{p} (X_{ik} - V_{kj})^2\right]^{\frac{-1}{w-1}}}{\sum_{j=1}^{p} \left[\sum_{j=1}^{p} (X_{ik} - V_{kj})^2\right]^{\frac{-1}{w-1}}} \quad (6)$$

6. Menghitung fungsi objektif pada iterasi ke-t menggunakan persamaan (6), dimana:
   P_t = Nilai fungsi objektif

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

c = Jumlah klaster
μ_{ik} = Derajat keanggotaan data ke-i terhadap klaster ke-k
X_{ij} = Nilai data ke-i pada dimensi j
V_{kj} = Pusat klaster ke-k pada dimensi j
d_{ik} = Jarak *Euclidean* antara data X_i dan pusat klaster V_k

$$P_t = \sum_{i=1}^{s}\sum_{k=i}^{c}\left(\left[\sum_{j=1}^{p} {d_{ik}}^2 (X_{ij}, V_{kj})^2\right](\mu_{ik})^w\right) \quad (6)$$

7. Mengecek kondisi berhenti:
   Jika $|P_t - P_{t-1}| < \xi$ atau $t > MaxIter$, maka berhenti
   Jika tidak: $t + 1$, ulangi langkah ke-4
8. Menentukan anggota pada setiap klaster
   Sebuah titik data dikatakan tergabung dalam suatu klaster jika nilai keanggotaannya mendekati 1 atau mencapai nilai terbesar

2.3 Koefisien *Silhouette*

Koefisien *silhouette* pertama kali dikembangkan oleh Peter Rousseeuw pada tahun 1987. Koefisien *Silhouette* menggabungkan *cohesion* (keseragaman dalam klaster) dan *separation* (pemisahan antara klaster) sebagai metode evaluasi untuk mengukur kualitas suatu klaster dan menentukan tingkat keanggotaan setiap objek dalam sebuah klaster [7][14]. Nilai koefisien *silhouette* berada pada kisaran -1 sampai dengan 1 yang menunju kkan jika semakin tinggi nilai koefisien *silhouette* (mendekati angka 1), maka semakin baik kualitas klaster tersebut. Sebaliknya jika semakin rendah nilai koefisien *silhouette* (mendekati angka -1), maka semakin buruk kualitas klaster tersebut [12][26]. Koefisien *silhouette* dihitung menggunakan persamaan (7), dimana:

SC = Nilai rata-rata *Silhouette Coefficient*
k = Jumlah klaster
n = Jumlah total data
$s(x_i)$ =Nilai *Silhouette Coefficient*
a_c =Rata-rata jarak dari x_i ke semua data lain dalam klaster yang sama
b_c = Rata-rata jarak dari x_i ke data dalam klaster terdekat

$$SC = \frac{1}{k}\Sigma_{i=1}^{n} S(x_c) \text{ dengan } s(x_i) = \frac{b_c - a_c}{max[b_c, a_c]} \quad (7)$$

2.4 *Davies-Bouldin Index*

Davies-Bouldin Index pertama kali diperkenalkan oleh David L. Davies dan Donald W. Bouldin pada tahun 1979 sebagai metode evaluasi untuk mengukur kualitas suatu klaster dengan memisahkan kelompok data yang berbeda dan mendekati pusat masing-masing klaster [27]. *Davies-Bouldin Index* merupakan rasio antara jarak rata-rata dalam klaster dengan jarak antar klaster tetangga terdekatnya. Prinsip dari *Davies-Bouldin Index* adalah untuk memaksimalkan jarak antar titik dalam klaster dan meminimalkan jarak antar klaster. Nilai *Davies-Bouldin Index* yang semakin tinggi menunjukkan semakin buruk kualitas klaster yang terbentuk. Sebaliknya jika nilai *Davies-Bouldin Index* yang semakin rendah menunjukkan semakin baik kualitas klaster yang terbentuk [28]. *Davies-Bouldin Index* dihitung menggunakan persamaan (8), dimana [28]:

R_{ij} = Rasio antara klaster i dan j
S_i = Jarak rata-rata semua titik dalam klaster i ke *centroid* klaster tersebut
S_j = Jarak rata-rata semua titik dalam klaster j ke *centroid* klaster tersebut
d_{ij} = Jarak antar *centroid* dari klaster i dan j

$$R_{ij} = \frac{S_i + S_j}{d_{ij}} \quad (8)$$

Langkah-langkah tahapan untuk menghitung *Davies-Bouldin Index* adalah sebagai berikut:

1. Tentukan posisi pusat klaster m_i dan m_j untuk setiap klaster C_i dan C_j
2. Hitung tingkat dispersi S_i untuk klaster C_i dan S_j untuk klaster C_j dengan caramenghitung jarak rata-rata antara titik data dalam klaster dan pusat klister tersebut.
3. Tentukan jarak d_{ij} antara pusat klaster m_i dan m_j
4. Gunakan rumus *Davies-Bouldin Index* untuk menghitung nilai R_{ij} antara klister C_i dan C_j
5. Ulangi kembali langkah 1 sampai 4 untuk setiap pasangan klaster yang ada

Setelah semua nilai R_{ij} dihitung, *Davies-Bouldin Index* dapat dihitung menggunakan persamaan (9), dimana [28]:
DB = *Davies-Bouldin Index* keseluruhan
N = Total klaster keseluruhan

$$DB = \frac{1}{n}\Sigma_{i=1}^{N} max_{j \neq I}(R_{ij}) \quad (9)$$

3.HASIL DAN PEMBAHASAN

Telah dilakukan pengujian terhadap dua algoritma *unsupervised learning* yang digunakan dalam penelitian ini untuk menganalisis klaster tindak pidana di Indonesia dengan menggunakan algoritma K-Means dan Fuzzy C-Means. Algoritma K-Means diujikan terlebih dahulu, kemudian dilanjutkan dengan pengujian menggunakan algoritma Fuzzy C-Means. Kedua algoritma itu diuji pada aplikasi *Visual Studio Code (VSC)* untuk menentukan algoritma terbaik dalam menganalisis klaster tindak pidana di Indonesia berdasarkan hasil dari nilai metode evaluasi yang digunakan yaitu Koefisien *silhouette* dan *Davies-Bouldin Index*. Hasil dari perbandingan algoritma K-Means dan Fuzzy C-Means dalam menganalisis klaster tindak pidana di Indonesia dapat dilihat pada Tabel (1) serta hasil output program pada Gambar (2).

Tabel 1 Perbandingan Nilai Evaluasi Antara Algoritma K-Means dan Fuzzy C-Means Dalam Menganalisis Klaster Tindak Pidana Di Indonesia

Algoritma	Hyperparameter	Eksperimen ke	Koefisien Silhouette	Davies-Bouldin Index
K-Means	n_clusters=2	1	0.7078	0.5457
		2	0.7078	0.5457
		3	0.7078	0.5457
		4	0.7078	0.5457
		5	0.7078	0.5457
		6	0.7078	0.5457
		7	0.7078	0.5457
		8	0.7078	0.5457
		9	0.7078	0.5457
		10	0.7078	0.5457
K-Means	n_clusters=3	1	0.5859	0.5481
		2	0.5859	0.5481
		3	0.5859	0.5481
		4	0.5859	0.5481
		5	0.5859	0.5481

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

Algoritma	Hyperparameter	Eksperimen ke	Koefisien Silhouette	Davies-Bouldin Index
		6	0.5859	0.5481
		7	0.5859	0.5481
		8	0.5859	0.5481
		9	0.5859	0.5481
		10	0.5859	0.5481
K-Means	n_clusters=4	1	0.4976	0.6009
		2	0.4976	0.6009
		3	0.4976	0.6009
		4	0.4976	0.6009
		5	0.4976	0.6009
		6	0.4976	0.6009
		7	0.4976	0.6009
		8	0.4976	0.6009
		9	0.4976	0.6009
		10	0.4976	0.6009
Fuzzy C-Means	n_clusters=2	1	0.7078	0.5457
		2	0.7078	0.5457
		3	0.7078	0.5457
		4	0.7078	0.5457
		5	0.7078	0.5457
		6	0.7078	0.5457
		7	0.7078	0.5457
		8	0.7078	0.5457
		9	0.7078	0.5457
		10	0.7078	0.5457
Fuzzy C-Means	n_clusters=3	1	0.4928	0.7468
		2	0.4928	0.7468
		3	0.4928	0.7468
		4	0.4928	0.7468
		5	0.4928	0.7468
		6	0.4928	0.7468
		7	0.4928	0.7468
		8	0.4928	0.7468
		9	0.4928	0.7468
		10	0.4928	0.7468
Fuzzy C-Means	n_clusters=4	1	0.4976	0.6009
		2	0.4976	0.6009
		3	0.4976	0.6009
		4	0.4976	0.6009
		5	0.4976	0.6009

Algoritma	Hyperparameter	Eksperimen ke	Koefisien Silhouette	Davies-Bouldin Index
		6	0.4976	0.6009
		7	0.4976	0.6009
		8	0.4976	0.6009
		9	0.4976	0.6009
		10	0.4976	0.6009

	Kepolisian Daerah	2000	2001	2002	2003	2004	2005	...	2017	2018	2019	2020	2021	2022	Cluster
1	ACEH	4286	3420	1668	2724.0	1873.0	2181.0	...	8885.0	8758	7483.0	7745.0	6651	10137.0	0
2	SUMATERA UTARA	15887	15395	15063	17530.0	20924.0	25111.0	...	39867.0	32922	30831.0	32990.0	36534	43555.0	1
3	SUMATERA BARAT	4464	4879	4845	5842.0	5387.0	7203.0	...	13205.0	12953	11064.0	7992.0	5666	7691.0	0
4	RIAU	4542	5341	5571	7020.0	7151.0	8839.0	...	10542.0	10655	9729.0	11037.0	9993	15747.0	0
5	JAMBI	1667	1493	1554	1793.0	1984.0	2202.0	...	9531.0	6313	6848.0	4709.0	3701	5359.0	0
6	SUMATERA SELATAN	10754	10152	10502	7534.0	7328.0	9797.0	...	17659.0	15606	14814.0	14120.0	14603	13525.0	0
7	BENGKULU	941	676	1170	1159.0	1086.0	1100.0	...	4867.0	3389	3453.0	3333.0	3493	3613.0	0
8	LAMPUNG	5473	5265	3290	3697.0	4624.0	4253.0	...	11089.0	8963	8534.0	7594.0	9764	11022.0	0
11	METRO JAYA	18098	33157	32649	37895.0	53404.0	57762.0	...	34767.0	34655	31934.0	26585.0	29103	32534.0	1
12	JAWA BARAT	17542	18561	18943	17188.0	18351.0	21520.0	...	28875.0	19832	16432.0	15506.0	10936	34523.0	1
13	JAWA TENGAH	12713	10978	10934	12528.0	13374.0	12823.0	...	12033.0	9127	10317.0	10712.0	8909	30060.0	0
14	DI YOGYAKARTA	1861	3077	3080	2063.0	2377.0	3429.0	...	7251.0	6731	6650.0	7721.0	4774	10591.0	0
15	JAWA TIMUR	25284	23608	25674	26347.0	25603.0	30476.0	...	34598.0	26295	26985.0	17642.0	19257	51905.0	1
17	BALI	6669	5980	4759	4354.0	5456.0	5902.0	...	3589.0	3212	3047.0	2597.0	2404	6304.0	0
18	NUSA TENGGARA BARAT	2904	3309	3109	3245.0	3429.0	4352.0	...	8132.0	6451	8185.0	8591.0	6296	5296.0	0
19	NUSA TENGGARA TIMUR	2634	3182	3165	3887.0	3468.0	5105.0	...	6729.0	6257	5865.0	4790.0	4909	5991.0	0
20	KALIMANTAN BARAT	1873	2239	2027	1846.0	2658.0	5145.0	...	6020.0	5814	4721.0	3858.0	4048	3975.0	0
21	KALIMANTAN TENGAH	2124	2111	1506	2230.0	2304.0	3026.0	...	2699.0	2667	2444.0	2629.0	2399	3189.0	0
22	KALIMANTAN SELATAN	3205	3872	3519	3542.0	3472.0	2757.0	...	6578.0	5699	5375.0	5206.0	4973	5016.0	0
23	KALIMANTAN TIMUR	4115	3908	3821	5264.0	5853.0	6778.0	...	9149.0	6683	5293.0	4624.0	5535	5501.0	0
25	SULAWESI UTARA	8943	10292	11932	11229.0	9934.0	11862.0	...	11080.0	13083	9792.0	8792.0	8660	12106.0	0
26	SULAWESI TENGAH	3787	4666	3480	3881.0	3246.0	4935.0	...	10240.0	9379	6265.0	5454.0	5139	5453.0	0
27	SULAWESI SELATAN	7660	7549	6740	7405.0	10033.0	12571.0	...	21616.0	21498	16008.0	12815.0	14636	28679.0	0
28	SULAWESI TENGGARA	2087	1483	1265	1183.0	1672.0	583.0	...	2866.0	1263	1213.0	2148.0	2431	3828.0	0
31	MALUKU	341	111	538	1851.0	1146.0	1444.0	...	3875.0	3473	4213.0	6200.0	4147	3603.0	0
34	PAPUA	2678	2522	3555	3694.0	4749.0	5387.0	...	6785.0	7311	6994.0	6962.0	6236	7584.0	0
30	SULAWESI BARAT	0	0	0	0.0	0.0	0.0	...	1841.0	1817	1863.0	1704.0	1500	2027.0	0
33	PAPUA BARAT	0	0	0	0.0	0.0	0.0	...	2284.0	3475	2972.0	3162.0	2784	4083.0	0

Gambar 2 Hasil Klastering Data Tindak Pidana Tahun 2000-2022 Di Indonesia

Dari hasil pengujian masing-masing algoritma pada Tabel (1), hasil klastering terbaik diraih saat algoritma K-Means dan Fuzzy C-Means menggunakan *hyperparameter* dengan *n_clusters=2*. Sedangkan hasil klastering terburuk diperoleh saat algoritma K-Means dan Fuzzy C-Means menggunakan *hyperparameter* dengan *n_clusters=4*. Pada Gambar (2), hasil *output* menunjukkan terdapat 2 kelompok tindak pidana dari 28 provinsi di Indonesia yang masing-masing provinsi dipilih sebagai bagian dari klaster 0 atau 1 menggunakan *hyperparameter* dengan *n_clusters=2*.

4.KESIMPULAN

Berdasarkan penelitian ini, kesimpulan yang diperoleh yaitu algoritma K-Means dan algoritma Fuzzy C-Means sama memperoleh hasil nilai metode evaluasi Koefisien *silhouette* dan *Davies-Bouldin Index* terbaik saat *hyperparameter* yang digunakan yaitu *n_clusters=2*. Masing-masing adalah 0.7078 untuk hasil nilai metode evaluasi Koefisien *silhouette* dan 0.5457 untuk hasil nilai metode evaluasi *Davies-Bouldin Index*. Akan tetapi jika *hyperparameter* yang digunakan yaitu *n_clusters=3*, maka algoritma K-Means jauh lebih baik dari segi nilai metode evaluasi Koefisien *silhouette* ataupun nilai metode evaluasi *Davies-Bouldin Index* daripada algoritma Fuzzy C-Means. Disimpulkan juga terdapat 24 provinsi Indonesia yang berada di klaster 0 dan 4 provinsi Indonesia yang berada di klaster 1, ini menunjukkan 24 provinsi tersebut memiliki rata-rata jumlah tindak pidana pertahun yang lebih sedikit ketimbang pada 4 provinsi tersebut yang memiliki rata-rata jumlah tindak pidana pertahun yang lebih tinggi. Oleh karena itu, pada 4 provinsi tersebut yaitu Sumatera Utara, Metro Jaya, Jawa Barat, dan Jawa Timur diperlukan strategi, pengawasan serta kebijakan-kebijakan pemerintah atau kepolisian yang lebih efektif guna menghindari terjadinya penambahan rata-rata jumlah tindak pidana pertahunnya pada 4 provinsi tersebut. Untuk pengembangan selanjutnya, diharapkan peneliti dapat memodifikasi kembali pada bagian *hyperparameter* algoritma atau menggunakan algoritma yang lain agar bisa mendapatkan hasil nilai metode evaluasi Koefisien *silhouette* yang lebih tinggi dan hasil nilai metode evaluasi *Davies-Bouldin Index* yang lebih rendah.

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

UCAPAN TERIMA KASIH

Terima kasih kepada kepolisian Indonesia yang telah mengumpulkan data jumlah tindak pidana di setiap provinsi di Indonesia dari tahun 2000 hingga tahun 2022. Terima kasih juga kepada Badan Pusat Statistik Indonesia yang telah berkontribusi dalam merilis data kepolisian tersebut melalui situs web, sehingga dapat diunduh oleh publik dan digunakan dalam penelitian ini.

DAFTAR PUSTAKA

[1] A. Hanan, A. F. N. Masruriyah, and T. A. Mudzakir, "Implementasi Model Prediksi Data Kriminalitas Menggunakan Algoritma Single Moving Average," *Scientific Student Journal for Information, Technology and Science*, vol. 5, no. 1, pp. 150–158, 2024.

[2] A. D. Sulistyono, A. N. Hanifah, I. Slamet, and A. Adnan, "Cluster Analysis using K-Means Algorithm and Centroid Linkage for Indonesia Crime Data 2021: Analisis Cluster Menggunakan Algoritma K-Means dan Centroid Linkage pada Data Kriminalitas Indonesia Tahun 2021," *Al-Musthalah: Jurnal Riset dan Penelitian Multidisiplin*, vol. 1, no. 1, pp. 54–67, 2021.

[3] E. Sarastuti, D. Mahdiana, and N. Kusumawardhany, "Klasterisasi Tindak Kriminalitas di Provinsi Jawa Barat dengan Menggunakan Algoritma K-Medoids," *Bit (Fakultas Teknologi Informasi Universitas Budi Luhur)*, vol. 21, no. 1, pp. 84–84, Apr. 2024, doi: https://doi.org/10.36080/bit.v21i1.2976.

[4] M. Julham, S. Sumarno, F. Anggraini, A. Wanto, and S. Solikhun, "Penerapan Jaringan Syaraf Tiruan dalam Memprediksi Tingkat Kriminal di Kabupaten Simalungun Menggunakan Algoritma Backpropagation," *Brahmana Jurnal Penerapan Kecerdasan Buatan*, vol. 1, no. 1, pp. 64–73, 2019, doi: https://doi.org/10.30645/brahmana.v1i1.9.

[5] F. Rahmadayanti and R. Rahayu, "PENERAPAN METODE DATA MINING PADA KASUS KRIMINALITAS INDONESIA," *Jurnal Teknologi Informasi MURA*, vol. 15, no. 1, pp. 52–61, Jun. 2023, doi: https://doi.org/10.32767/jti.v15i1.2054.

[6] R. Risdianti, A. K. Nasution, R. Oktaviandi, and E. Bu'ulolo, "Penerapan Algoritma Apriori Untuk Mengetahui Pola Jenis Kejahatan Yang Sering Terjadi (Studi Kasus: Polsek Percut Sei Tuan)," *Seminar Nasional Sains dan Teknologi Informasi (SENSASI)*, vol. 3, no. 1, pp. 117–120, 2021.

[7] R. N. Fahmi, M. Jajuli, and N. Sulistiyowati, "Analisis Pemetaan Tingkat Kriminalitas di Kabupaten Karawang menggunakan Algoritma K-Means," *INTECOMS: Journal of Information Technology and Computer Science*, vol. 4, no. 1, pp. 67–79, Jun. 2021, doi: https://doi.org/10.31539/intecoms.v4i1.2413.

[8] J. MacQueen, "Some methods for classification and analysis of multivariate observations," *Project Euclid*, vol. 5.1, pp. 281–298, 1967.

[9] "Steinhaus, H. (1957) Sur la division des corps matériels en parties. Bulletin L'Académie Polonaise des Science, 4, 801-804. (In French) - References - Scientific Research Publishing," *Scirp.org*, 2018.

[10] "Lloyd, S.P. (1957) Least Square Quantization in PCM. IEEE Transactions on Information Theory, 28, 129-137. - References - Scientific Research Publishing," *Scirp.org*, 2023.

[11] H. D. Tampubolon, S. Suhada, M. Safii, S. Solikhun, and D. Suhendro, "Penerapan Algoritma K-Means dan K-Medoids Clustering untuk Mengelompokkan Tindak Kriminalitas Berdasarkan Provinsi," *Jurnal Ilmu Komputer dan Teknologi*, vol. 2, no. 2, pp. 6–12, Nov. 2021, doi: https://doi.org/10.35960/ikomti.v2i2.703.

[12] N. Dinanti, S. Lestanti, and S. N. Budiman, "PENERAPAN ALGORITMA K – MEANS UNTUK PENGELOMPOKKAN TINDAK KEJAHATAN DI WILAYAH HUKUM POLRES BLITAR KOTA," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 5, pp. 3770–3777, Jan. 2024, doi: https://doi.org/10.36040/jati.v7i5.7815.

[13] D. A. Azhari, Y. Maulita, and S. Ramadani, "Pengelompokan Data Kriminal untuk Menentukan Pola Rawan Tindak Kriminal Menggunakan Algoritma K-Means," vol. 2, no. 5, pp. 122–138, Sep. 2024, doi: https://doi.org/10.62383/polygon.v2i5.238.

[14] W. W. Pribadi, A. Yunus, and A. S. Wiguna, "PERBANDINGAN METODE K-MEANS EUCLIDEAN DISTANCE DAN MANHATTAN DISTANCE PADA PENENTUAN ZONASI COVID-19 DI KABUPATEN MALANG," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 6, no. 2, pp. 493–500, Aug. 2022, doi: https://doi.org/10.36040/jati.v6i2.4808.

[15] D. Arthur and S. Vassilvitskii, "k-means++: the advantages of careful seeding," *Symposium on Discrete Algorithms*, pp. 1027–1035, Jan. 2007, doi: https://doi.org/10.5555/1283383.1283494.

[16] S. Vassilvitskii and D. Arthur, "K-means++: The Advantages of Careful Seeding." Accessed: Dec. 03, 2024. [Online]. Available: http://theory.stanford.edu/~sergei/slides/BATS-Means.pdf

[17] S. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, Mar. 1982, doi: https://doi.org/10.1109/tit.1982.1056489.

[18] Q. Du, V. Faber, and M. Gunzburger, "Centroidal Voronoi Tessellations: Applications and Algorithms," *SIAM Review*, vol. 41, no. 4, pp. 637–676, Jan. 1999, doi: https://doi.org/10.1137/s0036144599352836.

[19] "Elkan, C. (2003) Using the Triangle Inequality to Accelerate k-Means, In Proceedings of the Twentieth International Conference on International Conference on Machine Learning (ICML'03). AAAI Press, Washington DC, 147-153. - References - Scientific Research Publishing," *Scirp.org*, 2024.

[20] J. C. Dunn, "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters," *Journal of Cybernetics*, vol. 3, no. 3, pp. 32–57, Jan. 1973, doi: https://doi.org/10.1080/01969727308546046.

[21] A. Rohmatullah, D. Rahmalia, and M. S. Pradana, "KLASTERISASI DATA PERTANIAN DI KABUPATEN LAMONGAN MENGGUNAKAN ALGORITMA K-MEANS DAN FUZZY C MEANS," *Jurnal Ilmiah Teknosains*, vol. 5, no. 2, pp. 86–93, Feb. 2020, doi: https://doi.org/10.26877/jitek.v5i2.4254.

[22] G. N. S. Putri, D. Ispriyanti, and T. Widiharih, "IMPLEMENTASI ALGORITMA FUZZY C-MEANS DAN FUZZY POSSIBILISTICS C-MEANS UNTUK KLASTERISASI DATA TWEETS PADA AKUN TWITTER TOKOPEDIA," *Jurnal Gaussian*, vol. 11, no. 1, pp. 86–98, May 2022, doi: https://doi.org/10.14710/j.gauss.v11i1.33996.

[23] A. E. Pramitasari and Y. Nataliani, "PERBANDINGAN CLUSTERING KARYAWAN BERDASARKAN NILAI KINERJA DENGAN ALGORITMA K-MEANS DAN FUZZY C-MEANS," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 8, no. 3, pp. 1119–1132, Sep. 2021, doi: https://doi.org/10.35957/jatisi.v8i3.957.

[24] A. Muhammad and E. Budianita, "Pengelompokan Tingkat Kecanduan Game Online Menggunakan Algoritma Fuzzy C-Means," *Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)*, vol. 5, no. 4, pp. 601–610, Aug. 2022, doi: https://doi.org/10.32672/jnkti.v5i4.4511.

[25] N. Nurdiana, A. Nilogiri, and M. Rahman, "Penerapan Algoritma Fuzzy C-Means dan Metode Elbow untuk Mengelompokkan Provinsi di Indonesia Berdasarkan Indeks Demokrasi Indonesia," *Jurnal Smart Teknologi*, vol. 3, no. 5, pp. 544–551, 2022.

[26] M. D. Simatupang and A. W. Wijayanto, "ANALISIS KLASTER BERDASARKAN TINDAKAN KRIMINALITAS DI INDONESIA TAHUN 2019," *Jurnal Statistika Industri dan Komputasi*, vol. 6, no. 01, pp. 10–19, Jan. 2021.

[27] N. Azmi, H. HS, Y. Yuyun, and H. Hazriani, "Penerapan Metode K-Means Clustering Dalam Mengelompokkan Data Penjualan Obat pada Apotek M23," *Prosiding SISFOTEK*, vol. *7*, no. 01, pp. 244–248, Oct. 2023.

[28] I. T. Umagapi, B. Umaternate, H. Hazriani, and Y. Yuyun, "Uji Kinerja K-Means Clustering Menggunakan Davies-Bouldin Index Pada Pengelompokan Data Prestasi Siswa," *Prosiding SISFOTEK*, vol. 7, no. 1, pp. 303–308, Oct. 2023.