

Penerapan Algoritma K-Means Dan K-Medoids Clustering Untuk Mengelompokkan Tindak Kriminalitas Berdasarkan Provinsi

Abstract— *Crime is a problem that often occurs in everyday life and everywhere, including in various provinces in Indonesia. In this study, criminal acts will be grouped using the K-Means and K-Medoids clustering algorithms. The difference in the number of clusters in the performance of each algorithm has a different calculation pattern in each iteration depending on the dataset used and the centroid point that is used as a calculation in the algorithm. The data is processed into two clusters, namely the high crime rate cluster (C1) and the crime rate cluster. low crime (C2). The results of the K-Means algorithm show that C1 has 6 members and C2 has 28 members. While the results of the manual calculation of the K-Medoids algorithm obtained the results of C1 having 7 members and C2 having 27 members.*

Keywords— *Crime, K-Means, K-Medoids, Clustering*

Intisari— Kriminalitas merupakan maasalah yang sering terjadi di kehidupan sehari-hari dan dimana saja termasuk di berbagai provinsi yang ada di Indonesia. Pada penelitian ini akan dilakukan pengelompokkan tindak kriminalitas dengan menggunakan algoritma *K-Means* dan *K-Medoids clustering*. Perbedaan jumlah klaster pada kinerja tiap algoritma memiliki pola perhitungan yang berbeda pada masing-masing iterasi tergantung pada dataset yang digunakan serta titik centroid yang dijadikan perhitungan pada algoritma.. Data diolah menjadi dua *cluster* yaitu *cluster* tingkat tindak kriminalitas tinggi (C1) dan *cluster* tingkat tindak kriminalitas rendah (C2). Hasil algoritma *K-Means* diperoleh dengan C1 memiliki 6 anggota dan C2 memiliki 28 anggota. Sedangkan hasil perhitungan manual algoritma *K-Medoids* diperoleh hasil C1 memiliki 7 anggota dan C2 dengan memiliki 27 anggota.

Kata kunci — *klasifikasi, machine learning, algoritma C4.5, kelulusan uji kompetensi*

I. PENDAHULUAN

Kriminalitas merupakan kejahatan, tindak kriminal, atau juga sebagai suatu tindakan yang bersifat negatif. Seringkali, tindakan ini akan merugikan banyak pihak dan pelaku tindakannya disebut sebagai seorang kriminal. Kriminalitas sering terjadi di berbagai Negara, termasuk di Indonesia. Sebagian besar jenis kriminal yang sering terjadi di Indonesia adalah kejahatan terhadap nyawa (pembunuhan), kejahatan terhadap fisik (penganiayaan), kejahatan terhadap kesusilaan (pemeriksaan/pencabulan), pencurian, narkoba, penipuan bahkan korupsi [1].

Dengan banyaknya tindak kriminalitas di Indonesia, diperlukan adanya pengelompokan daerah rawan tindak kriminalitas di Indonesia berdasarkan provinsi sebagai salah satu usaha untuk membantu pihak kepolisian dalam mengambil keputusan apakah suatu daerah memerlukan pengawasan ekstra atau tidak. Oleh karena itu, maka dibuatlah pengelolaan data yang dapat menyelesaikan permasalahan tersebut yaitu pengelompokan daerah rawan

tindak kriminalitas di Indonesia menggunakan teknik data mining berupa analisis klaster (*clustering*). Di antara banyaknya analisis klaster yang ada terdapat dua jenis analisis klaster yang memiliki algoritma yang saling berkaitan, yaitu *K-Means* dan *K-Medoids clustering* [2].

Dalam penelitian yang dilakukan Rhamadani dan tim menggunakan metode *K-Means* dan *K-Medoids* untuk mengelompokkan nilai ujian nasional tingkat SMK menunjukkan bahwa pengelompokan data menggunakan algoritma *K-Means* dan *K-Medoids* memiliki pola perhitungan data yang berbeda, hasil pengelompokan pada algoritma *K-Means* menghasilkan klaster baik sebanyak 14 anggota, klaster sedang 46 anggota dan klaster cukup 49 anggota. Lalu, pada algoritma *K-Medoids* diperoleh hasil klaster baik 27 anggota, klaster sedang 48 anggota dan klaster cukup 43 anggota [3].

Penelitian yang dilakukan oleh Anggoro dan tim menggunakan metode *K-Means clustering*, *K-Medoids clustering* dan metode *Kernel Density* untuk mengelompokkan daerah rawan pencurian kendaraan bermotor (curanmor) di kota Semarang. Hasil pengujian yang didapat yaitu *K-Means* dan *K-Medoids* mempunyai hasil yang berbeda-beda. Sedangkan metode *Kernel Density* memiliki nilai verifikasi daerah rawan (tingkat kerawanan tinggi dan sedang) yang paling besar dibanding dengan metode *K-Means* dan *K-Medoids*, yaitu 82,35% [4].

Pada penelitian ini data kriminalitas yang digunakan diambil dari Badan Pusat Statistik melalui situs <https://www.bps.go.id/>. Data yang didapatkan akan dilakukan penelitian dalam mengelompokkan tindak kriminalitas berdasarkan provinsi menggunakan algoritma *K-Means* dan *K-Medoids*. Pengelompokan dibagi menjadi 2 *cluster* yaitu *cluster* tindak kriminalitas tinggi dan *cluster* tindak kriminalitas rendah. Pengelompokan dilakukan untuk menjadi sebuah solusi agar memudahkan pihak kepolisian dalam mengambil keputusan apakah suatu daerah memerlukan pengawasan ekstra atau tidak.

II. METODOLOGI PENELITIAN

A. Data Mining

Data mining adalah proses penggalian informasi dari data set yang besar melalui penggunaan algoritma dan teknik yang diambil dari bidang statistik, machine learning dan database sistem manajemen [5].

Data mining merupakan inti dari proses *Knowledge Discovery in Database* (KDD). KDD dapat dikatakan sebagai penyelesaian masalah dengan menganalisa data yang ada pada database dengan data tersimpan secara

elektronik dan pencariannya dilakukan otomatis seperti pada komputer [6].

B. Algoritma K-Means

K-Means merupakan salah satu algoritma *Clustering* yang masuk dalam kelompok *Unsupervised Learning* yang digunakan untuk membagi data menjadi beberapa kelompok dengan system partisi. Pada setiap *cluster* terdapat titik pusat (centroid) yang mempresentasikan *cluster* tersebut [7]. Prinsip utama dari algoritma *K-Means* ini adalah menyusun k buah partisi/pusat (centroid)/ rata-rata (mean) dari sekumpulan data.

Langkah-langkah algoritma *K-Means* [8] :

1. Tentukan jumlah *Cluster* (k) pada data set.
2. Tentukan titik pusat (Centroid) secara random.
3. Untuk menghitung jarak terdekat dengan centroid, dapat menggunakan rumus *Euclidean Distance* di bawah ini :
$$d = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \quad (1)$$

Keterangan :

d = *Euclidean Distance*
i = banyak objek
x, y = Titik koordinat objek
s, t = Titik koordinat centroid
4. Hitung pusat *cluster* yang baru dengan anggota *cluster* yang sekarang. Pusat *cluster* adalah rata-rata dalam sebuah *cluster*. Dapat dihitung menggunakan rumus:
$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj} \quad (2)$$

Keterangan:

V_{ij} = Centeroid rata-rata *cluster* ke-i untuk variable ke-j
 N_i = Jumlah anggota *cluster* ke-i
i, k = Indeks dari *cluster*
j = Indeks dari variable
 X_{kj} = Nilai data ke-k variable ke-j untuk *cluster* tersebut.
5. Lakukan kembali langkah ke-3 dan langkah ke-4 dan iterasi mencapai centroid bernilai optimal.

C. Algoritma K-Medoids

K-Medoids adalah salah satu metode partisi, karena menggunakan objek yang paling terpusat (medoids) di *cluster* menjadi pusat *cluster* dari nilai rata-rata objek dalam sebuah *cluster* [9].

Algoritma *K-Medoids* adalah algoritma *clustering* yang mirip dengan *K-Means*. Perbedaan dari kedua algoritma ini yaitu algoritma *K-Medoids* menggunakan objek sebagai perwakilan (medoid) sebagai pusat *cluster*, sedangkan *K-Means* menggunakan nilai rata-rata (mean) sebagai pusat *cluster*. Langkah-langkah algoritma *K-Medoids* [10] :

1. Inisialisasi pusat *cluster* sebanyak k (jumlah *cluster*).
2. Alokasikan setiap data (objek) ke *cluster* terdekat menggunakan persamaan ukuran jarak Euclidian Distance dengan persamaan (1).
3. Pilih secara acak objek pada masing-masing *cluster* sebagai kandidat medoids baru.

4. Hitung jarak setiap objek yang berada pada masing-masing *cluster* dengan kandidat medoids baru.
5. Hitung total simpangan (S) dengan menghitung nilai total distance baru – total distance lama. Jika $S < 0$, maka tukar objek dengan data *cluster* untuk membentuk sekumpulan k objek baru sebagai medoid.
6. Ulangi langkah 3 sampai 5 hingga tidak terjadi perubahan medoids, sehingga didapatkan *cluster* beserta anggota *cluster* masing-masing.

D. Pengujian RapidMiner

Rapidminer adalah perangkat lunak yang dapat diakses siapa saja dan bersifat terbuka (*open source*). *Rapidminer* memiliki kurang lebih 500 operator data mining, termasuk operator untuk *input*, *output*, data *preprocessing* dan visualisasi [11]. Pada penelitian ini, Pengujian data menggunakan tools *Rapidminer* 5.3 bertujuan untuk mencocokkan kesesuaian hasil dari perhitungan manual algoritma *K-Means* dan algoritma *K-Medoids*.

E. Tahapan Pengumpulan Data

Sumber data penelitian yang digunakan pada penelitian ini diambil dari website resmi Badan Pusat Statistik atau <https://www.bps.go.id/> yaitu tentang data kriminalitas menurut provinsi dari tahun 2017-2019 yang terdiri dari 34 provinsi. Variabel yang digunakan ada 6 yaitu tindak kejahatan terhadap nyawa (pembunuhan), kejahatan terhadap fisik (penganiayaan), kejahatan terhadap kesucilaan (pemerkosaan/pencabulan), kejahatan terhadap hak milik/barang (pencurian), narkoba, dan korupsi. Data akan diolah dengan melakukan *clustering* tindak kriminalitas menurut provinsi dalam 2 *cluster* yaitu *cluster* tindak kriminalitas tinggi (C1) dan *cluster* tindak kriminalitas rendah (C2).

F. Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) yaitu data tindak kriminalitas berdasarkan provinsi dari tahun 2017-2019 dengan jumlah data sebanyak 34 provinsi dengan kriteria tindak kejahatan terhadap nyawa, kejahatan terhadap fisik, kejahatan terhadap kesucilaan, kejahatan terhadap hak milik/barang, narkoba, serta penipuan dan korupsi. Berikut tabel data yang akan digunakan dengan nilai rata-rata (tahun 2017-2019) dari setiap kriteria:

TABEL 1. DATA TINDAK KRIMINALITAS BERDASARKAN PROVINSI

Provinsi	Tindak Kriminalitas					
	A	B	C	D	E	F
Aceh	25	1498	280	3063	1483	1310
Sumatera Utara	101	5230	254	11389	6123	4960
Sumatera Barat	15	2171	356	7382	735	1693
Riau	26	854	104	2497	1520	1004
Kepulauan Riau	7	601	128	1552	390	639
Jambi	21	775	91	3203	519	1343
Sumatera selatan	131	1835	188	4799	1797	2155
Kep. Bangka Belitung	12	119	44	1014	310	164
Bengkulu	29	406	90	1568	188	536
Lampung	27	446	178	2982	2009	1292

DKI Jakarta	62	1777	144	4872	7422	5603
Banten	16	337	126	1603	437	550
Jawa Barat	52	2002	445	6749	1236	4026
Jawa Tengah	33	661	241	4422	1216	1443
DI Yogyakarta	12	384	45	1828	274	924
Jawa timur	66	2363	259	7473	3252	3984
Bali	9	579	86	1296	729	440
Nusa Tenggara Barat	5	439	47	2096	236	662
Nusa Tenggara Timur	47	2422	220	2115	66	774
Kalimantan Barat	20	223	163	2222	644	577
Kalimantan Tengah	24	225	79	874	587	256
Kalimantan Selatan	35	371	57	3367	1529	450
Kalimantan Timur	17	658	117	1994	2238	639
Kalimantan Utara	13	77	40	520	404	298
Sulawesi Utara	74	3456	307	1908	381	1296
Gorontalo	14	1351	177	887	58	475
Sulawesi Tengah	15	1672	220	4665	296	1182
Sulawesi Barat	13	1099	56	811	299	292
Sulawesi Selatan	62	2817	222	6417	1015	3435
Sulawesi Tenggara	12	691	109	622	64	161
Maluku	11	793	141	857	103	316
Maluku Utara	11	189	122	123	79	47
Papua Barat	10	608	247	1317	56	361
Papua	44	1248	96	4135	32	561

Keterangan:

A = Kejahatan terhadap nyawa

B = Kejahatan terhadap fisik

C = Kejahatan terhadap kesusilaan

D = Kejahatan terhadap hak milik/barang

E = Narkoba

F = Penipuan dan korupsi

III. HASIL DAN PEMBAHASAN

A. Perhitungan Hasil Algoritma K-Means

1. Jumlah *cluster* yang digunakan adalah 2 *cluster*, yaitu *cluster* tingkat tindak kriminalitas tinggi (C1) dan *cluster* tingkat tindak kriminalitas rendah (C2).
2. Titik Centroid atau pusat *cluster* dipilih secara acak dari dataset. Pusat *cluster* tersebut dapat diketahui pada tabel 2 berikut:

TABEL 2. CENTROID DATA AWAL

Cluster	A	B	C	D	E	F
Sumatera Utara (C1)	101	5230	254	11389	6123	4960
Maluku Utara (C2)	11	189	122	123	79	47

3. Menghitung jarak setiap data terhadap centeroid (Pusat *Cluster*) menggunakan rumus Euclidean Distance pada persamaan 1 algoritma *K-Means*, seperti berikut ini:

$$\begin{aligned} & \text{Data}_{(1,1)} \\ &= \sqrt{(25 - 101)^2 + (1498 - 5230)^2 + (280 - 254)^2 + (3063 - 11389)^2 + (1483 - 6123)^2 + (1310 - 4960)^2} \\ &= 10867,78 \end{aligned}$$

$$\begin{aligned} & \text{Data}_{(2,1)} \\ &= \sqrt{(25 - 11)^2 + (1498 - 189)^2 + (280 - 122)^2 + (3063 - 123)^2 + (1483 - 79)^2 + (1310 - 47)^2} \\ &= 3734,78 \end{aligned}$$

$$\begin{aligned} & \cdot \\ & \cdot \\ & \cdot \\ & \text{Data}_{(1,34)} \\ &= \sqrt{(44 - 101)^2 + (1248 - 5230)^2 + (96 - 254)^2 + (4135 - 11389)^2 + (32 - 6123)^2 + (561 - 4960)^2} \\ &= 11178,39 \\ & \text{D}_{(2,34)} \\ &= \sqrt{(44 - 11)^2 + (1248 - 189)^2 + (96 - 122)^2 + (4135 - 123)^2 + (32 - 79)^2 + (561 - 47)^2} \\ &= 4181,60 \end{aligned}$$

Hasil perhitungan jarak data dengan titik pusat *cluster* pada iterasi 1 menggunakan *Euclidean distance* dapat dilihat pada tabel 3 berikut:

TABEL 3. HASIL PERHITUNGAN JARAK PUSAT CLUSTER ITERASI 1

Provinsi	Jarak ke Centroid		Jarak Terdekat
	Cluster 1	Cluster 2	
Aceh	10867,78	3734,78	C2
Sumatera Utara	0	14595,46	C1
Sumatera Barat	8070,63	7734,06	C2
Riau	11622,50	3011,80	C2
Kepulauan Riau	13029,09	1630,65	C2
Jambi	11462,02	3421,12	C2
Sumatera selatan	9030,04	5655,81	C2
Kep. Bangka Belitung	13806,06	933,76	C2
Bengkulu	13211,77	1545,14	C2
Lampung	11133,49	3676,71	C2
DKI Jakarta	7517,24	10481,72	C1
Banten	13096,16	1610,44	C2
Jawa Barat	7532,83	8029,19	C1
Jawa Tengah	10289,19	4686,19	C2
DI Yogyakarta	12862,66	1938,59	C2
Jawa timur	5722,86	9183,59	C1
Bali	13155,33	1451,30	C2
Nusa Tenggara Barat	12747,35	2088,96	C2
Nusa Tenggara Timur	12169,87	3081,22	C2
Kalimantan Barat	12583,67	2238,04	C2
Kalimantan Tengah	13726,86	932,23	C2
Kalimantan Selatan	11377,63	3581,39	C2
Kalimantan Timur	11956,55	2955,06	C2
Kalimantan Utara	14113,23	587,79	C2
Sulawesi Utara	11808,27	3943,21	C2
Gorontalo	13500,05	1456,22	C2
Sulawesi Tengah	10300,84	4916,70	C2
Sulawesi Barat	13591,004	1189,21	C2
Sulawesi Selatan	7678,74	7673,85	C2
Sulawesi Tenggara	14010,76	717,21	C2
Maluku Utara	14595,46	0	C2
Papua Barat	13445,28	1309,94	C2
Papua	11178,39	4181,60	C2

4. Menentukan centroid baru untuk iterasi selanjutnya dengan menghitung nilai rata-rata (mean) pada masing-masing *cluster*. Berikut adalah nilai centroid baru untuk iterasi 2:

TABEL 4. CENTROID DATA BARU

Cluster	A	B	C	D	E	F
C1	70,25	2843	275,5	7620,75	4508,25	4643,25

C2	26,33	966,83	145,9	2404,63	656,46	842,5
----	-------	--------	-------	---------	--------	-------

- Ulangi langkah 3-4 hingga posisi data sudah tidak mengalami perubahan. Dalam penelitian ini iterasi dilakukan sampai iterasi ke 3. Hasil akhir posisi *cluster* pada centroid iterasi ke 3 dapat dilihat pada tabel dibawah ini:

TABEL 5. PEMBAGIAN *CLUSTER* PADA ITERASI 3

Provinsi	Letak Cluster
Aceh	C2
Sumatera Utara	C1
Sumatera Barat	C1
Riau	C2
Kepulauan Riau	C2
Jambi	C2
Sumatera selatan	C2
Kep. Bangka Belitung	C2
Bengkulu	C2
Lampung	C2
DKI Jakarta	C1
Banten	C2
Jawa Barat	C1
Jawa Tengah	C2
DI Yogyakarta	C2
Jawa timur	C1
Bali	C2
Nusa Tenggara Barat	C2
Nusa Tenggara Timur	C2
Kalimantan Tengah	C2
Kalimantan Selatan	C2
Kalimantan Timur	C2
Kalimantan Utara	C2
Sulawesi Utara	C2
Gorontalo	C2
Sulawesi Tengah	C2
Sulawesi Barat	C2
Sulawesi Selatan	C1
Sulawesi Tenggara	C2
Maluku	C2
Maluku Utara	C2
Papua Barat	C2
Papua	C2

Pada Tabel 5 dijelaskan provinsi apa saja yang termasuk kategori tindak kriminalitas tinggi dan tindak kriminalitas rendah. Kategori tindak kriminalitas tinggi memiliki 6 anggota (Sumatera Utara, Sumatera Barat, DKI Jakarta, Jawa Barat, Jawa Timur, dan Sulawesi Selatan) dan tindak kriminalitas rendah memiliki 28 anggota (Aceh, Riau, Kepulauan Riau, Jambi, Sumatera Selatan, Kep. Bangka Belitung, Bengkulu, Lampung, Banten, Jawa Tengah, DI Yogyakarta, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Gorontalo, Sulawesi Tengah, Sulawesi Barat, Sulawesi Tenggara, Maluku, Maluku Utara, Papua Barat, dan Papua).

B. Perhitungan Hasil Algoritma K-Medoids

Tahap selanjutnya menerapkan dataset menggunakan algoritma *K-Medoids*. Tahapan penerapan algoritma *K-Medoids* hampir sama dengan tahapan pada algoritma *K-Means* yang membedakan hanya pada pemilihan titik centroid yang digunakan. *K-Medoids* melakukan pemilihan

titik centroid baru secara acak pada perhitungan jarak ke-2, sedangkan *K-Means* mengambil nilai mean pada jarak terkecil untuk dijadikan titik centroid di perhitungan jarak iterasi selanjutnya. Berikut adalah perhitungan algoritma *K-Medoids*:

- Inisialisasi pusat *cluster* sebanyak 2 *cluster* dari data sampel. Untuk pemilihan setiap *centroid* dipilih secara acak. Seperti pada tabel 6.

TABEL 6. MEDOID AWAL

Cluster	A	B	C	D	E	F
Sumatera Utara (C1)	101	5230	254	11389	6123	4960
Riau (C2)	26	854	104	2497	1520	1004

- Menghitung nilai jarak dengan persamaan Euclidean Distance yang ditunjukkan oleh persamaan 1 pada algoritma *K-Means*. Berikut tabel hasil perhitungan jarak pada iterasi 1 algoritma *K-Medoids*:

TABEL 7. HASIL PERHITUNGAN ALGORITMA *K-MEDOIDS* ITERASI KE-1

Provinsi	C1	C2	Jarak Terpendek	Cluster yang diikuti
Aceh	10867,78	927,94	927,94	C2
Sumatera Utara	0	11622,50	0	C1
Sumatera Barat	8070,63	5172,26	5172,26	C2
Riau	11622,50	0	0	C2
Kepulauan Riau	13029,09	1538,86	1538,86	C2
Jambi	11462,02	1273,49	1273,49	C2
Sumatera selatan	9030,04	2771,49	2771,49	C2
Kep. Bangka Belitung	13806,06	2216,53	2216,53	C2
Bengkulu	13211,77	1748,48	1748,48	C2
Lampung	11133,49	853,95	853,95	C2
DKI Jakarta	7517,24	7904,41	7517,24	C1
Banten	13096,16	1564,00	1564,00	C2
Jawa Barat	7532,83	5359,80	5359,80	C2
Jawa Tengah	10289,19	2011,67	2011,67	C2
DI Yogyakarta	12862,66	1493,67	1493,67	C2
Jawa timur	5722,86	6240,47	5722,86	C1
Bali	13155,33	1569,20	1569,20	C2
Nusa Tenggara Barat	12747,35	1449,94	1449,94	C2
Nusa Tenggara Timur	12169,87	2187,56	2187,56	C2
Kalimantan Barat	12583,67	1194,57	1194,57	C2
Kalimantan Tengah	13726,86	2111,96	2111,96	C2
Kalimantan Selatan	11377,63	1139,94	1139,94	C2
Kalimantan Timur	11956,55	969,75	969,75	C2
Kalimantan Utara	14113,23	2502,08	2502,08	C2
Sulawesi Utara	11808,27	2922,91	2922,91	C2
Gorontalo	13500,05	2293,87	2293,87	C2
Sulawesi Tengah	10300,84	2629,21	2629,21	C2
Sulawesi Barat	13591,00	2214,24	2214,24	C2
Sulawesi Selatan	7678,74	5039,81	5039,81	C2
Sulawesi Tenggara	14010,76	2524,48	2524,48	C2
Maluku	13727,27	2275,11	2275,11	C2
Papua Barat	13445,28	2007,57	2007,57	C2
Papua	11178,39	2291,08	2291,08	C2
Total Cost (Total Jarak Terdekat)			80507,46	

- Setelah didapatkan hasil jarak dari setiap objek pada iterasi 1 maka lanjut ke iterasi 2. Kandidat centroid baru pada iterasi 2 dapat dilihat pada tabel 8 berikut:

TABEL 8. KANDIDAT CENTROID BARU

Cluster	A	B	C	D	E	F
Sulawesi Selatan (C1)	62	2817	222	6417	1015	3435
Lampung (C2)	27	446	178	2982	2009	1292

Hasil perhitungan jarak titik centroid baru terhadap keseluruhan data proses seperti pada Tabel 9.

TABEL 9. HASIL PERHITUNGAN JARAK KE SETIAP NON MEDOIDS BARU

Provinsi	C1	C2	Jarak Terpendek	Cluster yang diikuti
Aceh	4210,51	1183,50	1183,50	C2
Sumatera Utara	7678,74	11133,49	7678,74	C1
Sumatera Barat	2116,99	4914,40	2116,99	C1
Riau	5039,81	853,95	853,95	C2
Kepulauan Riau	6066,21	2262,60	2262,60	C2
Jambi	4375,02	1545,11	1545,11	C2
Sumatera selatan	2416,21	2455,89	2416,21	C1
Kep. Bangka Belitung	6906,68	2856,06	2856,06	C2
Bengkulu	6199,38	2428,23	2428,23	C2
Lampung	4796,30	0	0	C2
DKI Jakarta	7016,02	7295,98	7016,02	C1
Banten	6163,89	2222,18	2222,18	C2
Jawa Barat	1105,62	4975,50	1105,62	C1
Jawa Tengah	3555,00	1665,97	1665,97	C2
DI Yogyakarta	5819,48	2121,11	2121,11	C2
Jawa timur	2574,53	5713,48	2574,53	C1
Bali	6348,73	2287,65	2287,65	C2
Nusa Tenggara Barat	5714,55	2084,01	2084,01	C2
Nusa Tenggara Timur	5161,87	2949,91	2949,91	C2
Kalimantan Barat	5712,95	1732,64	1732,64	C2
Kalimantan Tengah	6910,45	2756,39	2756,39	C2
Kalimantan Selatan	4948,51	1052,57	1052,57	C2
Kalimantan Timur	5792,28	1226,27	1226,27	C2
Kalimantan Utara	7247,88	3127,44	3127,44	C2
Sulawesi Utara	5071,88	3589,26	3589,26	C2
Gorontalo	6512,43	3111,60	3111,60	C2
Sulawesi Tengah	3158,43	2698,87	2698,87	C2
Sulawesi Barat	6693,26	3013,10	3013,10	C2
Sulawesi Selatan	0	4796,30	0	C1
Sulawesi Tenggara	7052,70	3270,59	3270,59	C2
Maluku	6751,24	3036,95	3036,95	C2
Maluku Utara	7673,85	3676,71	3676,71	C2
Papua Barat	6423,56	2735,77	2735,77	C2
Papua	4112,37	2534,26	2534,26	C2
Total Cost (Total Jarak Terpendek)			84930,98	

- Kemudian dilakukan perhitungan total simpangan (S) yang dihitung dari hasil perhitungan total cost (jarak terpendek) iterasi ke-1 dan ke-2. Dihitung dengan menggunakan perhitungan simpangan S dari jarak keduanya.

$$S = \text{total cost baru} - \text{total cost lama}$$

$$\text{Dengan : Total jarak iterasi ke-1} = 80507,46$$

$$\text{Total jarak iterasi ke-2} = 84930,98$$

$$\text{Total simpangan (S)} = 84930,98 - 80507,46 = 4423,52$$
- Karena nilai $S > 0$ maka proses pengklasteran dihentikan. Berikut hasil akhir dari pembagian anggota tiap klaster menggunakan algoritma *K-Medoids* dapat dilihat pada Tabel 10.

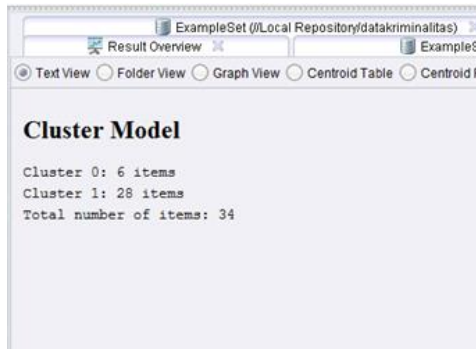
TABEL 10. PEMBAGIAN CLUSTER PADA ITERASI 2

Provinsi	Letak Cluster
Aceh	C2
Sumatera Utara	C1
Sumatera Barat	C1
Riau	C2
Kepulauan Riau	C2
Jambi	C2
Sumatera selatan	C1
Kep. Bangka Belitung	C2
Bengkulu	C2
Lampung	C2
DKI Jakarta	C1
Banten	C2
Jawa Barat	C1
Jawa Tengah	C2
DI Yogyakarta	C2
Jawa timur	C1
Bali	C2
Nusa Tenggara Barat	C2
Nusa Tenggara Timur	C2
Kalimantan Barat	C2
Kalimantan Tengah	C2
Kalimantan Selatan	C2
Kalimantan Timur	C2
Kalimantan Utara	C2
Sulawesi Utara	C2
Gorontalo	C2
Sulawesi Tengah	C2
Sulawesi Barat	C2
Sulawesi Selatan	C1
Sulawesi Tenggara	C2
Maluku	C2
Maluku Utara	C2
Papua Barat	C2
Papua	C2

Pada Tabel 10 dijelaskan provinsi apa saja yang termasuk kategori tindak kriminalitas tinggi dan tindak kriminalitas rendah. Kategori tindak kriminalitas tinggi memiliki 7 anggota (Sumatera Utara, Sumatera Barat, Sumatera Selatan, DKI Jakarta, Jawa Barat, Jawa Timur, dan Sulawesi Selatan) dan tindak kriminalitas rendah memiliki 27 anggota (Aceh, Riau, Kepulauan Riau, Jambi, Kep. Bangka Belitung, Bengkulu, Lampung, Banten, Jawa Tengah, DI Yogyakarta, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Gorontalo, Sulawesi Tengah, Sulawesi Barat, Sulawesi Tenggara, Maluku, Maluku Utara, Papua Barat, dan Papua).

C. Pengujian Menggunakan RapidMiner

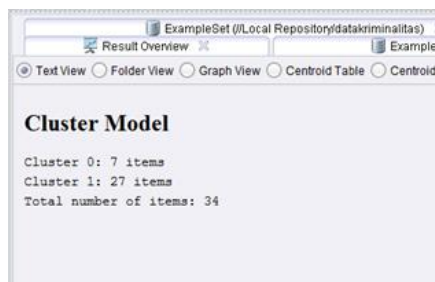
- Pada pengujian menggunakan Rapidminer algoritma *K-Means* menjelaskan bahwa hasil pengujian sesuai dengan perhitungan seperti pada gambar 1.



Gambar 1. Cluster Algoritma K-Means

Gambar 1 menjelaskan bahwa *cluster* 0 memiliki 6 anggota yang merupakan kategori tindak kriminalitas tinggi dan *cluster* 1 memiliki 28 anggota termasuk kedalam kategori tindak kriminalitas rendah dari hasil pengujian menggunakan algoritma *K-Means*.

- Pada pengujian menggunakan Rapidminer algoritma *K-Medoids* menjelaskan bahwa hasil pengujian sesuai dengan perhitungan seperti pada Gambar 2.



Gambar 2. Cluster Algoritma K-Medoids

Gambar 2 menjelaskan bahwa *cluster* 0 memiliki 7 anggota yang merupakan kategori tindak kriminalitas tinggi dan *cluster* 1 memiliki 27 anggota termasuk kedalam kategori tindak kriminalitas rendah dari hasil pengujian menggunakan algoritma *K-Means*.

IV. KESIMPULAN

Dari penelitian ini, dapat disimpulkan bahwa:

- Hasil algoritma *K-Means* diperoleh hasil *cluster* 1 dengan kategori tindak kriminalitas tinggi memiliki 6 anggota, *cluster* 2 dengan kategori tindak kriminalitas rendah memiliki 28 anggota. Sedangkan hasil perhitungan manual algoritma *K-Medoids* diperoleh hasil *cluster* 1 dengan kategori tindak kriminalitas tinggi memiliki 7 anggota, *cluster* 2 dengan kategori tindak kriminalitas rendah memiliki 27 anggota.
 - Hasil perhitungan manual algoritma *K-Means* diperoleh dengan menghitung sebanyak 3 iterasi, sedangkan perhitungan manual algoritma *K-Medoids* diperoleh sebanyak 2 iterasi.
- Pengujian data pada masing-masing algoritma terhadap tools menghasilkan kesesuaian jumlah *cluster* yang sama antara hasil perhitungan manual algoritma *K-Means* dengan Rapidminer dan algoritma *K-Medoids* dengan Rapidminer.

UCAPAN TERIMA KASIH

Terima kasih kepada program studi Teknik Informatika, STIKOM Tunas Bangsa, Pematangsiantar atas dukungan dalam terlaksananya penelitian ini serta pembimbing yaitu Bapak Suhada dan Bapak M. Safii.

REFERENCES

- [1] A. Supratman, "Menyoal Sikap Kejahatan Di Indonesia Di Era Industri 4.0 (Suatu Perspektif Kriminologi)," *Leg. J. Perundang Undangan dan Huk. Pidana Islam*, vol. 0, pp. 27–42, 2020.
- [2] Y. H. Susanti and E. Widodo, "Perbandingan K-Means dan K-Medoids Clustering terhadap Kelayakan Puskesmas di DIY Tahun 2015," *Pros. SI MaNIs (Seminar Nas. Integr. Mat. dan Nilai Islam.*, vol. 1, no. 1, pp. 116–122, 2017.
- [3] N. N. Rhamadani, A. Fauzi, E. Nurlaelasari, and A. R. Pratama, "Implementasi algoritma k-means dan k-medoids dalam pengelompokan nilai ujian nasional tingkat smk," *Conf. Innov. Appl. Sci. Technol. (CIASTECH 2020)*, no. Ciastech, pp. 717–726, 2020.
- [4] J. W. Anggoro, M. Awaluddin, and A. L. Nugraha, "Zonasi Daerah Rawan Pencurian Kendaraan Bermotor (Curamor) di Kota Semarang Dengan Menggunakan Metode Cluster Analysis," *Geod. Undip*, vol. 8, pp. 225–234, 2019.
- [5] U. T. Suryadi and Y. Supriatna, "Sistem Clustering Tindak Kejahatan Pencurian Di Wilayah Jawa Barat Menggunakan Algoritma K-Means," *J. Teknol. dan Komun. STMIK Subang*, vol. 12, no. 1, pp. 15–27, 2019, doi: 10.47561/a.v12i1.147.
- [6] R. T. Vlandari, "Pengelompokan Tingkat Keamanan Wilayah Jawa Tengah Berdasarkan Indeks Kejahatan Dan Jumlah Pos Keamanan Dengan Metode Klastering K-Means," *J. Ilm. SINUS*, vol. Vol 14, No, no. ISSN :1693-1173, pp. 59–72, 2016.
- [7] D. Gultom, H. D. Tampubolon, L. Y. Hutabarat, F. R. Ilmi H Zer, and D. Hartama, "Penerapan Algoritma K-Means Untuk Mengetahui Tingkat Tindak Kejahatan Daerah Pematangsiantar," *J. Teknol. Inf.*, vol. 4, no. 1, pp. 146–151, 2020.
- [8] S. Handi Kurniawan Sohdianata, "Penerapan Algoritma K-Means Untuk Menetapkan Kelompok Mutu Karet," *Progresif*, vol. 7, no. 2, pp. 2–9, 2011.
- [9] D. Marlina, N. Lina, A. Fernando, and A. Ramadhan, "Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak," *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 4, no. 2, p. 64, 2018, doi: 10.24014/coreit.v4i2.4498.
- [10] N. Pulungan, S. Suhada, and D. Suhendro, "Penerapan Algoritma K-Medoids Untuk Mengelompokkan Penduduk 15 Tahun Keatas Menurut Lapangan Pekerjaan Utama," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 329–334, 2019, doi: 10.30865/komik.v3i1.1609.

- [11] E. Elisa and A. Annurullah, "Data Mining dalam Menganalisis Faktor Alasan Pemilihan Perumahan," no. September, pp. 43–48, 2020.



Double-blind Peer Review Form

Instructions: The journal is committed to high academic standards, treating publication as a collaborative process between **Author**, **Reviewers** and **Editors**. The goal of the peer review process is to improve the academic and scientific quality of the submissions. **Reviewers** will work with the **Author** through a collaborative process to ensure academic and scientific integrity. Constructive criticism is a necessary part of this collaborative effort and as such shall be offered and received in a professional manner. **Reviewers are expected to return their reviews within 7 days, unless otherwise specified by the Editor.**

Thank you! Your help is highly appreciated.

Date						
Manuscript Title	Penerapan Algoritma K-Means Dan K-Medoids Clustering Untuk Mengelompokkan Tindak Kriminalitas Berdasarkan Provinsi					
Overview (Please rate each item by typing letter x in the appropriate box)						
		Poor	Below Average	Average	Good	Very good
1	The topic of this paper is relevant, timely, and of interest to the audience of this journal.			x		
2	This paper is based on rigorous academic standards.				x	
3	The content of this paper is technically accurate and sound.				x	
4	The abstract is concise and sufficient			x		
5	The introduction provides the necessary background information				x	
6	The research methodology for the study is appropriate and applied properly.				x	
7	The supporting evidence in this paper is strongly reliable and properly validated.				x	
8	The results of analysis are correctly interpreted and/or conclusions are sound.				x	
9	References are complete and appropriate.				x	
10	Tables and figures are appropriate and adequate.				x	
11	The paper is easy to read and free from grammatical or spelling errors.				x	

General Comments	
-	
Strengths of manuscript	
-	
Weakness of manuscript	
-	
Suggestions for improvement	
-	
Recommended disposition of the manuscript: check one. (type letter x in the appropriate box)	
Accept	x
Accept with minor revisions	
Accept subject to major revisions	
Invite resubmission for a new review after major revisions	
Reject	