

K-Means dan K-Medoids Clustering pada Kasus Tindak Pidana Khusus

K-Means and K-Medoids Clustering in Specific Criminal Cases

Ewing Rudita Arini*¹, M. Nur Haqqul Qomarudin², Yesinta Kharisma Ningtias³

^{1,2,3} Program Studi Matematika, Fakultas Ilmu Eksakta, Universitas Nahdlatul Ulama blitar

Correspondence author email: *ewingrarini@gmail.com

Abstrak

Tindak pidana khusus merupakan bagian dari hukum pidana yang diatur dalam undang-undang. Pada penelitian ini, metode clustering yang digunakan adalah K-Means dan K-Medoids dan dievaluasi menggunakan Davies Bouldin Index. Data yang digunakan pada penelitian ini adalah data tindak pidana khusus tahun 2023. Berdasarkan perhitungan diperoleh nilai DBI K-Means lebih rendah dibandingkan DBI pada K-Medoids. Untuk itu, metode K-Means dipilih untuk data tindak pidana khusus. Adapun hasil K-Means clustering adalah cluster-I sebanyak 8 wilayah diduga rawan tinggi, cluster-II sebanyak 11 wilayah diduga rawan sedang dan cluster-III sebanyak 14 wilayah diduga rawan rendah.

Kata kunci: K-Means, K-Medoids, Tindak Pidana Khusus

Abstrack

Specific criminal cases are part of criminal law regulated by law. In this research, the clustering methods are used K-Means and K-Medoids and evaluated using the Davies Bouldin Index. The data used in this research is specific criminal cases data in 2023. Based on the calculations obtained, the DBI value of K-Means is lower than the DBI value of K-Medoids. For this reason, the K-Means method was chosen for specific criminal cases. The results of K-Means clustering are that 8 areas suspected of being high risk, 11 areas suspected of being medium risk and 14 areas suspected of being low risk.

Keywords: K-Means, K-Medoids, Specific Criminal Cases

PENDAHULUAN

Tindak pidana adalah suatu istilah resmi yang digunakan dalam perundang-undangan pidana Indonesia. Tindak pidana adalah suatu perbuatan yang melanggar larangan hukum dan dapat dikenai sanksi pidana [5]. Berdasarkan data [3], ada kenaikan jumlah tindak pidana di Indonesia. Pada tahun 2022 ada sebanyak 372.897 dan tahun 2021 sebanyak 239.481. Artinya, ada kenaikan jumlah tindak pidana sebesar 133.416 untuk periode 2021-2022. Berdasarkan sumbernya, tindak pidana dapat dibedakan menjadi dua yaitu tindak pidana umum dan tindak pidana khusus. Tindak pidana umum adalah semua tindak pidana yang diatur dalam KUHP sebagai kodifikasi hukum pidana materil. Sedangkan, tindak pidana khusus adalah semua tindak pidana yang tidak tercakup dalam kodifikasi KUHP [10].

Tindak Pidana Khusus mendapat perhatian lebih dibandingkan tindak pidana lain di berbagai belahan dunia. Fenomena ini dapat dimaklumi mengingat dampak negatif yang ditimbulkannya. Dampaknya dapat mempengaruhi berbagai bidang kehidupan. Tindak pidana ini membahayakan stabilitas dan politik, serta dapat merusak moral dan nilai demokrasi karena perilaku tersebut lambat laun menjadi budaya. Bahkan setelah kemerdekaan, Indonesia masih menghadapi perang melawan tindak pidana.

Untuk memberantas tindak pidana khusus, pemerintah telah mengeluarkan beberapa peraturan. Dalam hal ini, Kejaksaan RI dan Polri berperan penting dalam memberantas tindak pidana khusus. Oleh karena itu, diperlukan suatu terobosan untuk memerangi kejahatan ini. Hal ini termasuk mendorong transparansi penyelenggaraan Negara serta merekomendasikan tindakan perbaikan kepada pemerintah dan memodernisasi layanan publik berbasis online dan menerapkan pengawasan terpadu.

Analisis klaster atau clustering adalah proses mempartisi sekumpulan data (atau pengamatan) ke dalam himpunan bagian (cluster/kelompok). Data dikelompokkan dengan memaksimalkan kesamaan intra-kelas dan meminimumkan kesamaan antar kelas. Data pada cluster yang sama mempunyai kemiripan yang tinggi tetapi berbeda dengan data pada cluster

lain. Metode pengelompokan yang berbeda pada kumpulan data yang sama akan menghasilkan pengelompokan yang berbeda pula. Oleh karena itu, analisis klaster berguna karena dapat mengarah pada penemuan kelompok yang sebelumnya tidak diketahui dalam data [6].

Analisis klaster sering digunakan untuk pengelompokan data ada dua jenis yaitu hierarchical dan non-hierarchical. Beberapa contoh metode data clustering non-hierarchical adalah K-Means dan K-Medoids. Algoritma K-Means mengelompokkan data dalam satu kelompok yang memiliki karakteristik yang sama serta memiliki karakteristik yang berbeda dengan data pada kelompok lain. K-Medoids adalah varian dari K-Means yang lebih kuat terhadap outlier. K-Medoids mengelompokkan partisi menggunakan titik aktual dalam cluster yang mewakilinya. Kelompok yang dibangun dihitung berdasarkan kedekatan yang dimiliki medoids dan obyek non-medoids [9].

Penelitian mengenai perbandingan algoritma K-Means dan K-Medoids sudah pernah dilakukan sebelumnya. Pada penelitian [8], waktu yang digunakan dalam memproses data transaksi bongkar muat menggunakan algoritma K-Means lebih cepat dibandingkan K-Medoids. Selain itu, nilai DBI pada K-Means lebih rendah dibandingkan nilai DBI K-Medoids. Penelitian lain yang dilakukan oleh [7], metode K-Means clustering memiliki hasil yang lebih baik dibandingkan metode K-Medoids Clustering karena menghasilkan nilai DBI yang lebih kecil. Data yang digunakan pada penelitian ini adalah data nilai yang diperoleh guru kelas pada enam bulan pertama awal pembelajaran pada sekolah luar biasa.

Berdasarkan uraian diatas, adapun tujuan dari penelitian ini ialah untuk membandingkan penerapan metode K-Means dan K-Medoids dalam pengelompokan wilayah hukum di Indonesia berdasarkan kasus tindak pidana khusus. Tindak pidana khusus yang dimaksud meliputi korupsi, narkoba, terorisme, dan kejahatan siber. Evaluasi cluster yang digunakan pada penelitian ini adalah Davies Bouldin Index. Penelitian ini diharapkan mampu membantu dan memberikan masukan kepada pemerintahan dalam membuat kebijakan serta merencanakan program dan evaluasi terhadap wilayah yang diduga rawan kasus tindak pidana khusus.

METODE PENELITIAN

Penelitian ini bertujuan untuk menerapkan algoritma K-Means *Clustering* dan K-Medoids *Clustering* serta membandingkan kedua algoritma tersebut menggunakan Davies Bouldin Index (DBI). Berikut ini adalah tahapan pada algoritma K-Means *Cluster*[1] :

- Tahap 1, menentukan banyak k-kelompok/*cluster* yang ingin dibentuk
- Tahap 2, menentukan pusat awal pada k-*cluster* secara acak
- Tahap 3, menghitung jarak setiap data ke setiap pusat *cluster* menggunakan *Euclidean distance*
- Tahap 4, mengelompokkan semua data berdasarkan jarak minimum ke pusat *cluster*
- Tahap 5, menentukan pusat *cluster* baru berdasarkan rata-rata semua data dalam *cluster*
- Tahap 6, mengulangi tahap 3 sampai 5 hingga pusat *cluster* tidak berubah atau tidak ada anggota *cluster* yang berpindah.

Sedangkan, tahapan pada K-Medoids *Clusterig* sebagai berikut [2]:

- Tahap 1, tentukan banyak k- kelompok yang ingin dibentuk
- Tahap 2, tentukan secara acak pusat awal pada k-*cluster*

- Tahap3, hitung jarak setiap data ke setiap pusat *cluster* dengan *Euclidean distance*
- Tahap 4, pilih secara acak kandidat/calon medoids baru di setiap *cluster*
- Tahap 5, hitung jarak setiap data di setiap *cluster* ke calon medoids baru
- Tahap 6, hitung total simpangan (S) dengan mengurangi nilai total jarak baru dengan nilai total jarak lama. Jika $S < 0$ maka tukarlah data tersebut dengan data *cluster* sehingga membentuk k objek baru sebagai medoid
- Tahap 7, ulangi tahap 4 sampai dengan 6 hingga tidak ada perubahan pada medoids

Metode yang dapat digunakan dalam evaluasi hasil *cluster* salah satunya adalah Davies Bouldin Index (DBI). Semakin rendah atau kecil nilai Davies Bouldin Index (DBI) maka semakin baik kualitas *cluster* yang diperoleh. Adapun metodenya sebagai berikut [4]:

1. SSW (*Sum of Square Within Cluster*) digunakan untuk mengetahui matrik kohesi pada *cluster* ke-i

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i) \quad (1)$$

2. SSB (*Sum of Square Between Cluster*) digunakan untuk mengetahui nilai separasi antara *cluster*

$$SSB_{i,j} = d(c_i, c_j) \quad (2)$$

3. Pengukuran rasio (R_{ij}) digunakan untuk mengetahui nilai perbandingan antar *cluster*

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (3)$$

4. Menghitung nilai DBI

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} R_{ij} \quad (4)$$

Keterangan

m_i = jumlah data pada *cluster* ke-i

$d(x_j, c_i)$ = jarak data ke-j ke centroid *cluster* ke-i

$d(c_i, c_j)$ = jarak antar centroid

k = jumlah *cluster*

HASIL DAN PEMBAHASAN

Penelitian ini menggunakan data sekunder yang diambil dari SIMKARI (Sistem Informasi Kejaksaan Republik Indonesia). Data yang digunakan adalah data tindak pidana khusus tahun 2023 untuk 33 wilayah di Indonesia.

Tabel 1. Data Tindak Pidana Khusus tahun 2023

Kantor Wilayah	Narapidana (X)	Tahanan (Y)
Aceh	4774	58
Bali	2288	38
Banten	4891	63
Bengkulu	1179	40
D. I. Yogyakarta	564	60
DKI Jakarta	8436	787
Gorontalo	128	40
Jambi	2618	47
Jawa Barat	10647	296
Jawa Tengah	6201	141
Jawa Timur	12885	133
Kalimantan Barat	3113	57
Kalimantan Selatan	6930	25
Kalimantan Tengah	2281	36
Kalimantan Timur	7859	69
Kepulauan Bangka Belitung	1191	34
Kepulauan Riau	2754	38
Lampung	3869	58
Maluku	233	73
Maluku Utara	296	36
Nusa Tenggara Barat	1827	56
Nusa Tenggara Timur	243	108
Papua	926	14
Papua Barat	430	32
Riau	7736	101
Sulawesi Barat	564	30
Sulawesi Selatan	3600	188
Sulawesi Tengah	1563	62
Sulawesi Tenggara	1262	58
Sulawesi Utara	176	49
Sumatera Barat	3148	61
Sumatera Selatan	7881	55
Sumatera Utara	15703	139

Tahap pertama dalam algoritma *K-Means* adalah menentukan jumlah *cluster*. Jumlah *cluster* yang akan digunakan pada penelitian ini adalah tiga. Tahap kedua adalah menentukan secara acak pusat *cluster* awal, yaitu

Tabel 2. Pusat *Cluster* Awal

Pusat <i>Cluster</i> ke-	Data ke-	X	Y
I	1	4774	58
II	14	2281	36
III	29	1262	58

Tahap ketiga dan keempat adalah menghitung jarak masing-masing data ke tiap pusat *cluster* dan mengelompokkan data berdasarkan jarak minimum dengan pusat *cluster*.

Tabel 3. Pengelompokan Data

Data ke-	C1	C2	C3	Cluster
1	0.00	2493.10	3512.00	C1
2	2486.08	7.28	1026.19	C2
3	117.11	2610.14	3629.00	C1
4	3595.05	1102.01	84.93	C3
5	4210.00	1717.17	698.00	C3
6	3733.86	6200.65	7210.94	C1
7	4646.03	2153.00	1134.14	C3
8	2156.03	337.18	1356.04	C2
9	5877.82	8370.04	9388.02	C1
10	1429.41	3921.41	4939.70	C1
11	8111.35	10604.44	11623.24	C1
12	1661.00	832.26	1851.00	C2
13	2156.25	4649.01	5668.10	C1
14	2493.10	0.00	1019.24	C2
15	3085.02	5578.10	6597.01	C1
16	3583.08	1090.00	74.95	C3
17	2020.10	473.00	1492.13	C2
18	905.00	1588.15	2607.00	C1
19	4541.02	2048.33	1029.11	C3
20	4478.05	1985.00	966.25	C3
21	2947.00	454.44	565.00	C2
22	4531.28	2039.27	1020.23	C3
23	3848.25	1355.18	338.87	C3
24	4344.08	1851.00	832.41	C3
25	2962.31	5455.39	6474.14	C1
26	4210.09	1717.01	698.56	C3
27	1181.18	1327.73	2341.61	C1
28	3211.00	718.47	301.03	C3

29	3512.00	1019.24	0.00	C3
30	4598.01	2105.04	1086.04	C3
31	1626.00	867.36	1886.00	C2
32	3107.00	5600.03	6619.00	C1
33	10929.30	13422.40	14441.23	C1

Pada tabel 3 diperoleh jumlah anggota *cluster* 1, 2 dan 3 berturut-turut 13,7 dan 13. Tahap kelima adalah menentukan pusat *cluster* baru dengan menghitung rata-rata semua data pada *cluster* yang sama, dan diperoleh

Tabel 4. Pusat *Cluster* Hasil Iterasi Pertama

Pusat <i>Cluster</i> ke-	X	Y
I	7800.923	162.538
II	2575.571	47.571
III	673.462	48.923

Tahap keenam adalah mengulangi tahap ketiga sampai kelima hingga pusat *cluster* tidak berubah atau tidak ada anggota *cluster* yang berpindah. Hasil perhitungan K-Means berhenti pada iterasi keempat.

Tabel 5. Pusat Custer dan Anggota *Cluster* Iterasi Terakhir

<i>Cluster</i> ke-	Pusat <i>Cluster</i>		Anggota <i>Cluster</i>
	X	Y	
I	9759.625	200.625	6,9,11,13,15,25,32,33
II	3594.273	71.364	1,2,3,8,10,12,14,17,18,27,31
III	755.857	49.428	4,5,7,16,19,20,21,22,23,24,26,28,29,30

Tahap selanjutnya adalah menghitung menggunakan algoritma K-Medoids. Tahap pertama sampai ketiga sama pengerjaannya seperti pada perhitungan K-Means. Perbedaannya terletak pada tahap keempat yaitu menghitung total cost jarak terdekat.

Tabel 6. Total Cost

Data ke-	C1	C2	C3	Cost	<i>Cluster</i>
1	0.00	2493.10	3512.00	0.00	C1
2	2486.08	7.28	1026.19	7.28	C2
3	117.11	2610.14	3629.00	117.11	C1
4	3595.05	1102.01	84.93	84.93	C3
5	4210.00	1717.17	698.00	698.00	C3
6	3733.86	6200.65	7210.94	3733.86	C1
7	4646.03	2153.00	1134.14	1134.14	C3
8	2156.03	337.18	1356.04	337.18	C2
9	5877.82	8370.04	9388.02	5877.82	C1
10	1429.41	3921.41	4939.70	1429.41	C1
11	8111.35	10604.44	11623.24	8111.35	C1

12	1661.00	832.26	1851.00	832.26	C2
13	2156.25	4649.01	5668.10	2156.25	C1
14	2493.10	0.00	1019.24	0.00	C2
15	3085.02	5578.10	6597.01	3085.02	C1
16	3583.08	1090.00	74.95	74.95	C3
17	2020.10	473.00	1492.13	473.00	C2
18	905.00	1588.15	2607.00	905.00	C1
19	4541.02	2048.33	1029.11	1029.11	C3
20	4478.05	1985.00	966.25	966.25	C3
21	2947.00	454.44	565.00	454.44	C2
22	4531.28	2039.27	1020.23	1020.23	C3
23	3848.25	1355.18	338.87	338.87	C3
24	4344.08	1851.00	832.41	832.41	C3
25	2962.31	5455.39	6474.14	2962.31	C1
26	4210.09	1717.01	698.56	698.56	C3
27	1181.18	1327.73	2341.61	1181.18	C1
28	3211.00	718.47	301.03	301.03	C3
29	3512.00	1019.24	0.00	0.00	C3
30	4598.01	2105.04	1086.04	1086.04	C3
31	1626.00	867.36	1886.00	867.36	C2
32	3107.00	5600.03	6619.00	3107.00	C1
33	10929.30	13422.40	14441.23	10929.30	C1
Total cost				54831.641	

Tahap kelima adalah melanjutkan perhitungan dengan menentukan secara acak pusat *cluster* baru dengan memilih data ke-1, 2 dan 3 sebagai medoids baru. Ulangi tahap kedua sampai kelima dan diperoleh total cost baru 666368.372. Selanjutnya, tahap ke-enam adalah menghitung total simpangan (S) yaitu menghitung selisih nilai total jarak baru dengan total jarak lama.

$$\begin{aligned}
 S_{total\ cost} &= Total\ Cost\ baru - Total\ Cost\ lama \\
 &= 666368.372 - 54831.641 \\
 &= 111536.731 > 0
 \end{aligned}$$

Karena nilai $S_{total\ cost}$ yang diperoleh bernilai positif (>0) maka proses iterasi berhenti, dan diperoleh hasil akhir pengelompokan

Tabel 7. Anggota Cluster Pada Iterasi Terakhir

Cluster ke-	Jumlah Anggota Cluster	Anggota Cluster
I	3	1,18,27
II	20	2,4,5,7,8,12,14,16,17,19,20,21,22,23,24,26,28,29,30,31
III	10	3,6,9,10,11,13,15,25,32,33

Algoritma K-Means dan K-Medoids selanjutnya dihitung menggunakan DBI (Davies Bouldin Index) untuk mengevaluasi manakah algoritma yang lebih rendah (atau lebih baik). Adapun langkahnya sebagai berikut

1. Langkah pertama, menghitung (SSW)

Tabel 8. Menghitung SSW

<i>Cluster ke-</i>	K-Means	K-Medoids
I	2507.571	695.392
II	985.626	1197.089
III	484.272	4034.041

2. Langkah kedua, menghitung SSB

Tabel 9. Menghitung SSB

	K-Means			K-Medoids		
	<i>Cluster-I</i>	<i>Cluster-II</i>	<i>Cluster-III</i>	<i>Cluster-I</i>	<i>Cluster-II</i>	<i>Cluster-III</i>
<i>Cluster-I</i>	0	6166.707	9005.037	0	2486.080	117.107
<i>Cluster-II</i>	6166.707	0	2838.500	2486.080	0	2603.120
<i>Cluster-III</i>	9005.037	2838.500	0	117.107	2603.120	0

3. Langkah ketiga, menghitung rasio (R_{ij})

Tabel 10. Menghitung Rasio

	K-Means			K-Medoids		
	<i>Cluster-I</i>	<i>Cluster-II</i>	<i>Cluster-III</i>	<i>Cluster-I</i>	<i>Cluster-II</i>	<i>Cluster-III</i>
<i>Cluster-I</i>	0	0.566	0.333	0	0.761	40.386
<i>Cluster-II</i>	0.566	0	0.519	0.761	0	2.010
<i>Cluster-III</i>	0.333	0.519	0	40.386	2.010	0

4. Langkah keempat, menghitung nilai DBI

$$DBI_{K-Means} = \frac{1}{3} (0.566 + 0.566 + 0.519) = 0.551$$

$$DBI_{K-Medoids} = \frac{1}{3} (40.386 + 2.010 + 40.386) = 27.594$$

Semakin kecil nilai DBI maka semakin baik kualitas *cluster* yang diperoleh. Berdasarkan perhitungan diperoleh bahwa nilai DBI pada K-Means lebih rendah atau lebih kecil dibandingkan DBI K-Medoids. Dengan demikian, metode K-Means dinilai cocok digunakan untuk data pada penelitian ini.

KESIMPULAN

Nilai DBI K-Means lebih rendah dari DBI K-Medoids, sehingga data tindak pidana khusus tahun 2023 lebih baik menggunakan metode K-Means. Adapun hasil K-Means clustering adalah cluster-I sebanyak 8 wilayah diduga rawan tinggi, cluster-II sebanyak 11 wilayah diduga rawan sedang dan cluster-III sebanyak 14 wilayah diduga rawan rendah. Metode clustering yang digunakan pada penelitian ini adalah K-Means dan K-Medoids. Untuk itu, penelitian selanjutnya diharapkan dapat menggunakan metode clustering lain agar mampu memberikan informasi yang berguna bagi masyarakat dan pemerintah terkait.

DAFTAR PUSTAKA

- [1]. Arini, E., “Penerapan K-Means Cluster Di Provinsi Jawa Timur Berdasarkan Ketahanan Pangan”, *Journal of Science Nusantara*, vol. 3, No. 1, pp. 32-36, 2023, doi: <https://doi.org/10.28926/jsnu.v3i1.866>
- [2]. Azra, M., *Penggunaan Davies Bouldin Index Dalam Perbandingan Algoritma K-Means dan K-Medoids untuk Klasterisasi Provinsi di Indonesia Berdasarkan Tingkat Indikator Ketenagakerjaan*, Skripsi, Fakultas Sains dan Teknologi, Universitas Jambi, 2023, Jambi
- [3]. Badan Pusat Statistik, “Jumlah Tindak Pidana Menurut Kepolisian Daerah, 2021-2022”, <https://www.bps.go.id/id/statistics-table/2/MTAxIzI=/jumlah-tindak-pidana-menurut-kepolisian-daerah.html>, diakses tanggal 03 Mei 2024 pukul 10.00
- [4]. Biantara, B., dkk, “Perbandingan Algoritma K-Means dan DBSCAN untuk Pengelompokan Data Penyebaran Covid-19 Seluruh Kecamatan di Provinsi Jawa Barat”, *Scientific Student Journal for Information, Technology and Science*, vol. IV, No. 1, pp. 88-94, 2023, <http://journal.ubpkarawang.ac.id/mahasiswa/index.php/ssj/article/view/722>
- [5]. Chandra, T., “Hukum Pidana”, 2022, PT. Sangir Multi Usaha, Jakarta
- [6]. Fitriana, R., dkk, “Data Mining dan Aplikasinya Contoh Kasus di Industri Manufaktur dan Jasa”, 2022, Wawasan Ilmu, Banyumas
- [7]. Harahap, F., “Perbandingan Algoritma K Means dan K Medoids Untuk Clustering Kelas Siswa Tunagrahita”, *TIN: Terapan Informatika Nusantara*, vol. 2, No. 4, pp: 191-197, 2021, <https://ejurnal.seminar-id.com/index.php/tin/article/view/873>
- [8]. Kamila, I., dkk, “Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau”, *Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi*, vol. 5, No. 1, pp. 119-125, 2019, doi: 10.24014/rmsi.v5i1.7381
- [9]. Wanto, A., dkk, “Data Mining : Algoritma dan Implementasi”, 2020, Yayasan Kita Menulis, Medan
- [10]. Wahyuni, F., “Dasar-dasar Hukum Pidana di Indonesia”, 2017, PT Nusantara Persada Utama, Tangerang Selatan