

SISTEM CLUSTERING DAERAH RAWAN KEMATIAN ANAK DI BAWAH UMUR DI WILAYAH JAWA BARAT MENGGUNAKAN ALGORITMA K-MEANS

Raditya Aji Sasmoyo

Teknik Informatika, Universitas Singaperbangsa Karawang
Jalan Raya HS.Ronggo Waluyo, Telukjambe Timur, Karawang, Indonesia
asviewers25@gmail.com

ABSTRAK

Kematian anak di bawah umur, apakah disebabkan oleh tanggung jawab langsung orang tua atau faktor lainnya, tetap menjadi tanggung jawab utama orang tua karena anak-anak masih tergolong rentan dan belum mampu melindungi diri sendiri sepenuhnya. Berdasarkan rilis tahunan Dinas Kesehatan pada tahun 2019, Provinsi Jawa Barat menunjukkan angka kematian bayi dan balita yang cukup tinggi, berada setelah Banten dan Kalimantan Timur. Pengelompokan wilayah yang dilakukan oleh Dinas Kesehatan bertujuan untuk memetakan daerah rawan kematian, sebagai upaya membantu pemerintah dalam menekan angka kematian balita, bayi, dan neonatal di Indonesia serta mempermudah penanganan penyebab kematian tersebut. Penelitian ini menggunakan algoritma K-Means untuk mengelompokkan daerah rawan kematian berdasarkan data Dinas Kesehatan tahun 2018-2019. Hasil pengkategorian membagi daerah menjadi tiga kategori (*cluster*), yaitu cluster 2 mencakup 1 kabupaten, cluster 1 mencakup 21 kabupaten/kota, dan cluster 0 mencakup 5 kabupaten/kota. Evaluasi pengelompokan dihasilkan nilai yaitu 0,62 menggunakan *silhouette coefficient*, hasil ini berarti kriteria yang ditunjukkan pengelompokan yang dilakukan memiliki struktur cluster yang cukup baik atau *medium structure*.

Kata kunci : Kematian anak di bawah umur, K-means clustering, Silhouette coefficient

1. PENDAHULUAN

Rawannya kematian terjadi pada anak di bawah umur mencakup anak balita, bayi bahkan neonatal (belum lahir) sangat mengkhawatirkan. Apalagi hal-hal yang menyebabkan kematian anak di bawah itu ditanggung jawab penuh oleh orang tua maka bisa dihindari. Agar mencegah kesedihan atas gagalannya orangtua dalam melindungi anak di bawah umur mereka.

Berdasarkan permasalahan tersebut di atas, penelitian ini bertujuan untuk mengklasifikasikan daerah rawan kematian pada anak kecil, bayi dan neonatus [1]. Dengan menggunakan teknik clustering yang merupakan salah satu teknik data mining. Penelitian ini memakai algoritma k-means dalam melihat pola jenis kematian yang akan membentuk kelompok regional. Hasil yang akan diperoleh berupa klaster wilayah terdapat kasus kematian pada balita, bayi dan neonatus di Indonesia. Manfaat penelitian ini adalah membantu Pemerintah dan Dinas Kesehatan (Dinkes) Jawa Barat untuk menurunkan angka kematian anak balita, bayi dan neonatus serta memudahkan penanganannya secara merata.

Kinerja algoritma K-Means dianggap optimal dalam teknik pengelompokan (*clustering*) karena menghasilkan pengelompokan data yang lebih akurat. Algoritma ini banyak digunakan karena sederhana dan efisien [2]. *K-Means* juga cenderung lebih scalable dan efisien dalam memproses objek dalam jumlah besar karena memiliki akurasi yang optimal untuk ukuran objek [3]. Meskipun memiliki kelebihan, algoritma K-Means juga memiliki kelemahan dalam menentukan jumlah cluster yang optimal untuk pengelompokan data dalam sebuah dataset [4]. Untuk mengatasi masalah ini, metode siku (*elbow method*) sering

digunakan untuk mengoptimalkan jumlah cluster dan mencari nilai cluster terbaik [5]. Metode *elbow* menentukan jumlah cluster terbaik dengan melihat perbandingan antara penurunan *Sum of Square Error* tiap cluster pada titik tertentu [5].

2. TINJAUAN PUSTAKA

2.1. Data Mining

Data mining adalah disiplin ilmu yang ditemukannya pola dari suatu data atau dipelajarinya metode untuk mengekstrak pengetahuan [6]. Data mining, yang biasa dikenal sebagai KDD (*knowledge discovery in database*), mencakup kegiatan seperti pengumpulan dan untuk ditemukannya pola, hubungan, atau keteraturan dalam set data yang besar dengan penggunaan data historis. Hasil dari data mining ini dapat digunakan untuk memperbaiki pengambilan keputusan di masa depan. Secara umum, data mining memiliki empat tugas utama:

a. Klasifikasi (*Classification*)

Klasifikasi memiliki tujuan untuk mengelompokkan item data ke dalam suatu kelas dari beberapa kelas yang telah ditentukan. Contohnya, sebuah program email dapat mengelompokkan email menjadi email sah dan email spam. Beberapa algoritma yang digunakan dalam klasifikasi antara lain pohon keputusan, nearest neighbor, naïve bayes, neural networks, dan support vector machines.

b. Regresi (*Regression*)

Regresi adalah pemodelan dan investigasi hubungan antara dua atau lebih variabel. Dalam analisis regresi, terdapat satu atau lebih variabel independen atau prediktor yang biasanya dilambangkan dengan notasi x dan satu variabel

- respon yang biasanya dilambangkan dengan notasi y .
- c. Pengelompokan (Clustering)
Clustering merupakan pendekatan untuk mengumpulkan serangkaian data ke dalam cluster (kelompok) di mana masing-masing cluster mengandung data yang serupa atau memiliki karakteristik yang mirip.
 - d. Pembelajaran Aturan Asosiasi (*Association Rule Learning*)
Pembelajaran aturan asosiasi berfokus pada penemuan keterhubungan antar variabel. Sebagai contoh, toko dapat mengumpulkan data tentang kebiasaan belanja pelanggan. Dengan menggunakan teknik ini, toko dapat mengungkap pola pembelian di mana produk tertentu sering dibeli bersama, yang kemudian dapat dimanfaatkan untuk strategi pemasaran yang lebih efektif.

2.2. Metode Clustering

Suatu metode salah satunya adalah clustering untuk mengidentifikasi sekelompok data dengan memisahkan mereka berdasarkan karakter-karakter tertentu yang sebelumnya ditentukan [6]. Dalam proses pengelompokan ini, label nilai belum ditetapkan, sehingga setelah dilakukan pengelompokan, diharapkan data tersebut dari tabel dapat diidentifikasi. Metode clustering sering dianggap sebagai langkah awal sebelum menerapkan metode lain, seperti klasifikasi.

Cluster analysis adalah proses mengelompokkan objek data berdasarkan kemiripan atau kesamaan informasi antara satu dengan yang lainnya. Tujuannya adalah untuk menemukan kelompok yang berkualitas, di mana objek-objek dalam satu kelompok memiliki kemiripan atau hubungan satu sama lain, dan berbeda dengan objek-objek dalam kelompok lainnya.

Clustering cocok digunakan untuk mengeksplorasi data. Jika terdapat banyak kasus tetapi tidak ada pengelompokan yang jelas, algoritma clustering dapat digunakan untuk menemukan pola dalam data tersebut. Selain itu, clustering juga bermanfaat sebagai tahap awal dalam pemrosesan data (data preprocessing), yaitu langkah untuk mengidentifikasi kelompok-kelompok yang berkaitan dalam rangka membangun model.

2.3. Algoritma K-Means

Algoritma K-Means clustering dikenal sebagai metode klasterisasi berbasis jarak yang bertujuan untuk data dibagi ke dalam beberapa kelompok [6]. Pendekatan ini mengelompokkan data berdasarkan ciri-ciri yang masing-masing data miliki. Data dengan ciri-ciri yang serupa dikelompokkan bersama dalam satu kelompok, sedangkan data dengan ciri-ciri yang berbeda dikelompokkan dalam kelompok yang berbeda.

Dalam algoritma ini, pusat dari setiap kelompok disebut sebagai centroid. Pada awalnya, posisi centroid

dipilih secara acak dari seluruh data yang ada. Selanjutnya, K-Means mengelompokkan masing-masing data di dalam kategori yang sesuai dengan centroid terdekat, berdasarkan jarak antara data dan centroid tersebut. Proses ini berulang hingga tidak ada lagi perubahan kelompok untuk setiap data.

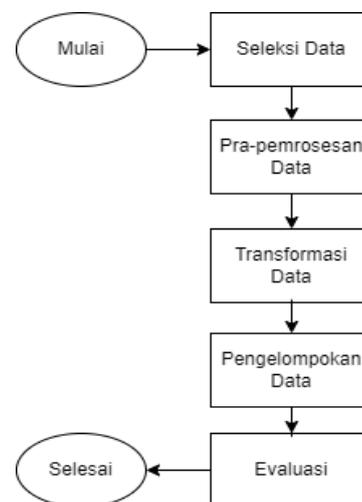
2.4. Penelitian Sebelumnya

Penelitian terkait pemakaian clustering dengan algoritma K-Means telah dilakukan oleh Johan Oscar Ong (2013), Muhammad Toha dkk. (2013), Ari Muzakir (2014), Sylvia Pretty Tulus (2014), serta Nurhayati dan Luigi Ajeng Pratiwi (2015). Dalam penelitiannya, Johan Oscar Ong (2013) mengumpulkan semua data dan kemudian menginisialisasikan data tersebut dalam bentuk angka agar dapat diproses dengan memakai metode K-Means clustering [7].

Dalam penelitiannya, Muhammad Toha dkk. (2013) melakukan pengkategorian siswa didasarkan karakter para siswa. Dalam penelitian ini, siswa dikategorikan menjadi empat cluster yaitu kelompok siswa berkarakter lemah, mulai terlihat, berkembang, dan kelompok siswa berkarakter unggul [8]. Ari Muzakir (2014) menentukan penerimaan beasiswa berdasarkan nilai Komputer, Matematika, dan Bahasa Inggris menggunakan algoritma K-Means dengan tiga klaster. Menghasilkan nilai yang memenuhi kriteria dalam kategori baik, yang dalam pembahasan ini adalah nilai di atas 70 [9].

Sylvia Pretty Tulus (2014), dalam penelitiannya, data spasial dikelompokkan menggunakan algoritma K-Means setelah melalui proses normalisasi. Dalam proses ini, pengelompokan dilakukan berdasarkan jarak terdekat antar objek, bukan berdasarkan karakteristik objek tersebut. [10]. Nurhayati dan Luigi Ajeng Pratiwi (2015) menggunakan dua cluster untuk mengkategorikan jurusan siswa yang diberi label cluster yaitu IPS dan IPA [11].

3. METODE PENELITIAN



Gambar 1. Diagram alur metodologi penelitian

Bagian ini menjelaskan metodologi penelitian yang digunakan. Proses penelitian terdiri dari lima langkah yang dijalankan secara berurutan. Langkah-langkah ini meliputi pemilihan data, pra-pemrosesan data, transformasi data, pengelompokan data, dan evaluasi hasil pengelompokan. Diagram alur metodologi penelitian sebagaimana ditunjukkan dalam gambar 1.

3.1. Pemilihan Data

Langkah pertama adalah proses pemilihan data yang akan digunakan dalam proses pengelompokan

data. Penelitian ini menggunakan dataset yang diunduh dari situs publik web terbuka data.jabarprov.go.id yang dikeluarkan oleh Dinas Kesehatan Jawa Barat. Data yang diperoleh adalah jumlah kematian menurut jenis kematian di Jawa Barat tahun 2018-2019. Dataset terdiri dari 299 catatan dengan 10 atribut. Atribut tersebut mencakup id, kode wilayah, nama wilayah, kode kabupaten/kota, nama kabupaten/kota, jenis kematian, dan jumlah kematian. Isi dataset awal ditunjukkan dalam tabel 1.

Tabel 1. Dataset awal tahun 2018-2019

Id	Kode	Provinsi	Kode Kota	Kabupaten/Kota	Jenis Kematian	Jenis Kelamin	Jumlah	Satuan	Tahun
1	32	JAWA BARAT	3201	KABUPATEN BOGOR	ANAK BALITA	LAKI-LAKI	0	JIWA	2018
2	32	JAWA BARAT	3201	KABUPATEN BOGOR	ANAK BALITA	PEREMPUAN	1	JIWA	2018
3	32	JAWA BARAT	3201	KABUPATEN BOGOR	BAYI	LAKI-LAKI	59	JIWA	2018
4	32	JAWA BARAT	3201	KABUPATEN BOGOR	BAYI	PEREMPUAN	50	JIWA	2018
5	32	JAWA BARAT	3201	KABUPATEN BOGOR	NEONATAL	LAKI-LAKI	51	JIWA	2018
...
295	32	JAWA BARAT	3279	KOTA BANJAR	ANAK BALITA	PEREMPUAN	5	JIWA	2019
296	32	JAWA BARAT	3279	KOTA BANJAR	BAYI	LAKI-LAKI	13	JIWA	2019
297	32	JAWA BARAT	3279	KOTA BANJAR	BAYI	PEREMPUAN	12	JIWA	2019
298	32	JAWA BARAT	3279	KOTA BANJAR	NEONATAL	LAKI-LAKI	10	JIWA	2019
299	32	JAWA BARAT	3279	KOTA BANJAR	NEONATAL	PEREMPUAN	7	JIWA	2019

3.2. Pra-Pemrosesan Data

Dalam penelitian ini, tahap pra-pemrosesan data yang dilakukan adalah menghapus atribut-atribut yang tidak diperlukan agar data siap digunakan pada tahap selanjutnya. Dalam penelitian ini, atribut yang tetap digunakan adalah nama kabupaten/kota, jenis kematian, dan jumlah kematian. Hasil pemrosesan pra-pemrosesan data setelah menghapus atribut yang tidak diperlukan sebagaimana ditunjukkan dalam tabel 2.

Tabel 2. Data setelah dilakukan pra-pemrosesan data

Kabupaten/Kota	Jenis Kematian	Jumlah
KABUPATEN BOGOR	ANAK BALITA	0
KABUPATEN BOGOR	ANAK BALITA	1
KABUPATEN BOGOR	BAYI	59
KABUPATEN BOGOR	BAYI	50
KABUPATEN BOGOR	NEONATAL	51
...
KOTA BANJAR	ANAK BALITA	5

Kabupaten/Kota	Jenis Kematian	Jumlah
KOTA BANJAR	BAYI	13
KOTA BANJAR	BAYI	12
KOTA BANJAR	NEONATAL	10
KOTA BANJAR	NEONATAL	7

3.3. Transformasi Data

Proses transformasi data dilakukan dengan mengubah struktur data. Pada dataset awal, terdapat dua kolom yang menunjukkan jenis kematian dan jumlah kematian. Kemudian, kolom-kolom ini diubah menjadi tiga kolom, masing-masing untuk jenis kematian dan jumlah kematian. Pada tahap ini, juga ditambahkan kolom baru yaitu jumlah atribut kematian yang menunjukkan jumlah kematian dari masing-masing kabupaten/kota. Hasil dari transformasi data sebagaimana ditunjukkan dalam tabel 3.

Tabel 3. Dataset dilakukan transformasi data

Kabupaten/Kota	Jenis Kematian Anak Balita	Jenis Kematian Bayi	Jenis Kematian Neonatal	Total Kematian
KABUPATEN BOGOR	7	213	185	405
KABUPATEN SUKABUMI	0	0	0	0

Kabupaten/Kota	Jenis Kematian Anak Balita	Jenis Kematian Bayi	Jenis Kematian Neonatal	Total Kematian
KABUPATEN CIANJUR	3	79	70	152
KABUPATEN BANDUNG	15	421	379	815
KABUPATEN GARUT	2	271	255	528
KABUPATEN TASIKMALAYA	33	442	343	818
KABUPATEN CIAMIS	29	207	156	392
KABUPATEN KUNINGAN	5	147	117	269
KABUPATEN CIREBON	15	269	243	527
KABUPATEN MAJALENGKA	4	83	68	155
KABUPATEN SUMEDANG	27	304	202	533
KABUPATEN INDRAMAYU	21	488	447	956
KABUPATEN SUBANG	5	225	210	440
KABUPATEN PURWAKARTA	17	134	101	252
KABUPATEN KARAWANG	9	319	274	602
KABUPATEN BEKASI	2	76	76	154
KABUPATEN BANDUNG BARAT	4	144	137	285
KABUPATEN PANGANDARAN	8	66	47	121
KOTA BOGOR	12	93	92	197
KOTA SUKABUMI	4	35	27	66
KOTA BANDUNG	27	227	163	417
KOTA CIREBON	6	37	29	72
KOTA BEKASI	4	111	97	212
KOTA DEPOK	8	148	120	276
KOTA CIMAHI	4	114	84	202
KOTA TASIKMALAYA	22	142	104	268
KOTA BANJAR	14	25	17	56

3.4. Pengelompokan Data

Dalam penelitian ini akan dilakukan pengelompokan (clustering) yang menggunakan algoritma k-means. Secara umum, algoritma K-means dijalankan dengan langkah-langkah sebagai berikut [12]:

- Menentukan n atau Jumlah dari cluster.
- Menentukan titik pusat (centroid) cluster yang dapat dilakukan secara acak.
- Jarak setiap data terhadap centroidnya masing-masing dilakukan penghitungan. Untuk penghitungan jarak ini dipakai rumus menggunakan rumus jarak Euclidean (Euclidean Distance) dengan persamaan (1) sebagai berikut:

$$d(i, j) = \sqrt{\sum_{k=1}^n (X_{ij} - C_{kj})^2} \quad (1)$$

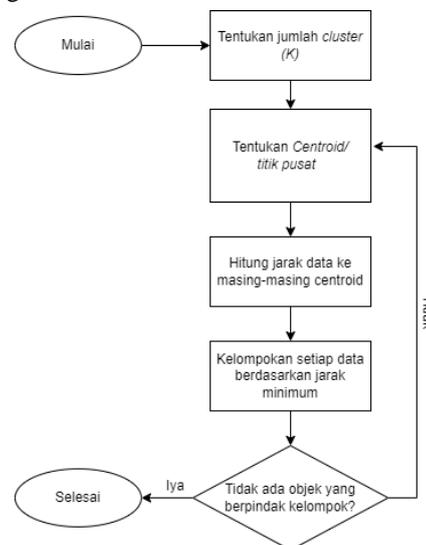
Keterangan:
 d(X, Y) : Jarak objek antara objek i dan j
 n : banyaknya dimensi atau atribut data
 x_{ik} : nilai pusat dari objek i pada dimensi k
 y_{jk} : nilai pusat cluster dari objek j pada dimensi k
- Mengategorikan data dengan didasarkan jarak terdekat antara centroid dengan data.
- Nilai centroid yang baru akan ditentukan yaitu dengan nilai rata-rata dari cluster yang disangkutkan, dengan persamaan (2) sebagai berikut:

$$C_k = \frac{1}{n_k} \sum d_i \quad (2)$$

Keterangan:

- n_k* : jumlah data dalam cluster k
 - d_i* : jumlah dari nilai jarak yang masuk dalam masing-masing cluster
- f. Langkah 3-5 di atas dilakukan perulangan sampai tiap cluster, anggotanya tidak ada perubahan.

Aluran implementasi algoritma k-means menggunakan flowchart, sebagaimana ditunjukkan dalam gambar 2.



Gambar 2. Flowchart Algoritma K-Means

3.5. Evaluasi

Evaluasi dilakukan untuk mengevaluasi kualitas cluster yang terbentuk pada tahap pengelompokan. Pengujian dalam penelitian ini menggunakan metode silhouette coefficient. Silhouette coefficient menggabungkan metode validasi clustering, yaitu cohesion yang menilai kedekatan antar objek dalam sebuah cluster, dan separation yang menilai jarak antara cluster [13]. Jika nilai silhouette coefficient mendekati atau positif, maka struktur cluster dikategorikan sebagai baik. Sebaliknya, jika nilai silhouette coefficient = 0, maka struktur cluster dianggap tidak jelas. Pengkategorian nilai dari silhouette coefficient sebagaimana ditunjukkan dalam tabel 4 berikut ini [14].

Tabel 4. Kategori nilai silhouette coefficient

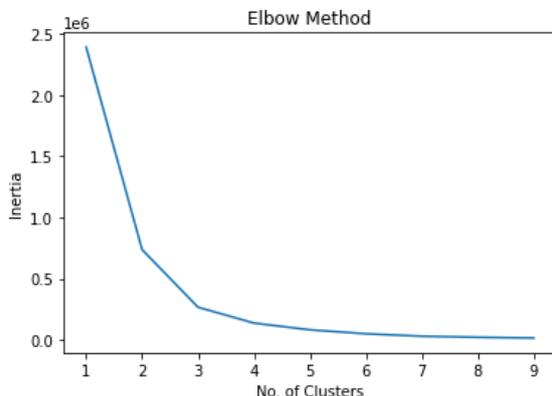
Nilai SC	Kriteria
0,71 – 1,00	Struktur cluster kuat (Strong Structure)
0,51 – 0,70	Struktur cluster standar (Medium Structure)
0,26 – 0,50	Struktur cluster lemah (Weak Structure)
≤ - 0,25	Tidak memiliki struktur (No Structure)

4. HASIL DAN PEMBAHASAN

Diterapkannya algoritma K-means dalam penelitian ini menggunakan bahasa pemrograman Python untuk melakukan clustering. Tahap awal proses pengelompokan melibatkan penentuan jumlah cluster yang optimal menggunakan metode elbow. Metode elbow ini digunakan untuk menentukan jumlah cluster dengan menganalisis persentase dari membandingkan antar jumlah cluster yang akan membentuk siku pada titik tertentu dengan nilai Sum of Square Error dari setiap cluster. Grafik metode elbow sebagaimana ditunjukkan dalam gambar 3.

```

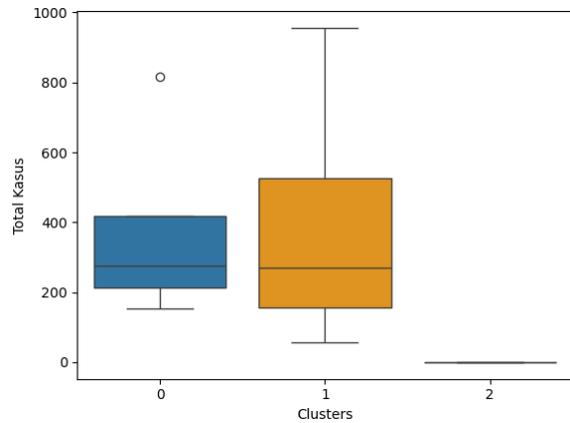
K : 1 Inertia : 2391726.5185185187
K : 2 Inertia : 737328.8684210527
K : 3 Inertia : 267485.95833333334
K : 4 Inertia : 137911.08333333334
K : 5 Inertia : 82935.83333333334
K : 6 Inertia : 50818.361904761914
K : 7 Inertia : 30780.02857142857
K : 8 Inertia : 22290.43333333334
K : 9 Inertia : 16001.600000000002
    
```



Gambar 3. Grafik Metode Elbow

Grafik diatas memperlihatkan dibentuknya 3 kategori antara lain cluster 0 (C0), cluster 1 (C1),

cluster 2 (C2). Untuk kemudian pengambilan jarak terpendek dari masing-masing data dilakukan agar dapat mengelompokkan. Hasil pengelompokkan ini sebagaimana ditunjukkan dalam tabel 5.



Gambar 5. Data sesudah pengkategorian cluster

Tabel 5 menunjukkan cluster 0 (C0) meliputi 5 kabupaten/kota meliputi Kabupaten Bandung, Kabupaten Bekasi, Kota Bandung, Kota Bekasi, dan Kota Depok. Cluster 1 (C1) meliputi 21 kabupaten/kota antara lain, Kabupaten Cianjur, Kabupaten Garut, Kabupaten Ciamis, Kabupaten Bogor, Kabupaten Tasikmalaya, Kabupaten Cirebon, Kabupaten Kuningan, Kabupaten Sumedang, Kabupaten Majalengka, Kabupaten Subang, Kabupaten Indramayu, Kabupaten Karawang, Kabupaten Purwakarta, Kabupaten Pangandaran, Kabupaten Bandung Barat, Kota Sukabumi, Kota Bogor, Kota Cimahi, Kota Cirebon, Kota Banjar, dan Kota Tasikmalaya. Sedangkan cluster 2 (C2) meliputi 1 kabupaten yaitu Kabupaten Sukabumi. Digunakannya silhouette coefficient menunjukkan evaluasi pengelompokan didapatkan nilai 0,6248969309270467 yang termasuk dalam struktur cluster standar (medium structure) berdasarkan kriteria pengelompokan yang dilakukan.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa pengelompokan wilayah menggunakan algoritma k-means menghasilkan pembagian menjadi 3 cluster berdasarkan karakteristik jenis kematian anak, bayi, dan neonatal di Jawa Barat. Ketiga cluster tersebut adalah: cluster 0 (C0) yang terdiri atas 5 kota/kabupaten, cluster 1 (C1) yang terdiri atas 21 kota/kabupaten, dan cluster 2 (C2) yang terdiri atas 1 kota/kabupaten. Evaluasi pengelompokan dihasilkan nilai 0,6248969309270467 dengan memakai silhouette coefficient, menunjukkan bahwa kriteria pengelompokan termasuk dalam struktur klaster standar (medium structure).

Saran untuk penelitian selanjutnya adalah menggunakan lebih banyak data untuk meningkatkan nilai indeks dalam proses evaluasi pemodelan. Selain itu, melakukan eksperimen dengan parameter yang

berbeda dapat membantu mengetahui perbandingan evaluasi yang lebih optimal.

DAFTAR PUSTAKA

- [1] R. Rahma, R. Mufidah, "Pengelompokan Daerah Rawan Kekerasan Terhadap Perempuan Dan Anak Di Jawa Barat Menggunakan Algoritma K-Means," *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 07, no. 3, pp. 850–857, 2022.
- [2] W. Sirait, S. Defit, dan G. W. Nurcahyo, "Algoritma K-Means Untuk Klasterisasi Tugas Akhir Mahasiswa Berdasarkan Keahlian," *J. Sistim Inf. dan Teknol.*, vol. 1, no. 3, pp. 25–30, 2019.
- [3] A. Bastian, H. Sujadi, dan G. Febrianto, "Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka)," *Jurnal Sistem Informasi (Journal of Information System)*, vol. 14, no. 1, pp. 26–32.
- [4] N. Putu, E. Merliana, dan A. J. Santoso, "Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means," pp. 978–979.
- [5] I. Wahyudi, M. B. Sulthan, dan L. Suhartini, "Analisa Penentuan Cluster Terbaik Pada Metode K-Means Menggunakan Elbow Terhadap Sentra Industri Produksi Di Pamekasan," *J. Apl. Teknol. Inf. dan Manaj.*, vol. 2, no. 2, pp. 72–81, 2021.
- [6] Prasetyo, Eko. *Data Mining: Konsep dan Aplikasi Menggunakan Matlab*. Jakarta: Andi Publisher, 2012.
- [7] J. O. Ong, "Implementasi Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing President University," *Jurnal Ilmiah Teknik Industri*, vol. 12, no. 1, pp. 10–20, 2013.
- [8] R. Toha, K. E. Purnama, dan S. Sumpeno, "Clustering Pencapaian Karakter Siswa Menggunakan Algoritma K-Means" *Prosiding Seminar Nasional Manajemen Teknologi XVII*, 2013.
- [9] Muzakir, A, "Analisa Dan Pemanfaatan Algoritma K-Means Clustering Pada Data Nilai Siswa Sebagai Penentuan Penerima Beasiswa," *Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST)*, pp. 1-6, 2014.
- [10] S. P. Tulus, "Perancangan Clustering Data Menggunakan Algoritma K-Means Berbasis Heat Map," 2014.
- [11] Nurhayati, & Luigi Ajeng Pratiwi, "Penerapan Algoritma K-Means dalam Data Mining untuk Peminatan Jurusan Bagi Siswa Kelas X (Studi Kasus: SMA Negeri 29 Jakarta)," *Prosiding Ilmiah Nasional Teknologi Komputer (SENATKOM 2015)*, vol 1. Padang : Universitas Putra Indonesia YPTK, 2015.
- [12] A. T. Rahman, Wiranto, dan A. Rini, "Coal Trade Data Clustering Using K-Means (Case Study Pt. Global Bangkit Utama)," *ITSMART J. Teknol. dan Inf.*, vol. 6, no. 1, pp. 24–31, 2017.
- [13] U. T. Suryadi dan 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.
- [14] Rousseeuw, P. J. 1987. "Silhouettes : A Graphical Aid To The Interpretation And Validation of Cluster Analysis," *Journal of Computational And Applied Mathematics*, 20, 53-56.