



Optimasi Algoritma *K-Means* Menggunakan Metode *Elbow* dalam Pengelompokan Penyakit Demam Berdarah Dengue (DBD) di Jawa Barat

Dea Amelia¹, Tesa Nur Padilah², Asep Jamaludin³

^{1,2,3}Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang

Abstract

Received: 22 Juni 2022

Revised: 25 Juni 2022

Accepted: 30 Juni 2022

Dengue hemorrhagic fever (DHF) is an acute febrile infectious disease that usually occurs in tropical and subtropical areas of the world and is caused by a virus transmitted by the Aedes mosquito, namely Aedes aegypti and Aedes albopictus. Dengue fever is one of the endemic diseases that almost occurs throughout the world. Indonesia is the country with the highest dengue fever cases in Southeast Asia. One of the provinces with the most cases of dengue fever is West Java. Every year cases of dengue fever have increased and decreased, so cases cannot be controlled properly. This must be a concern for the West Java Government in handling this Dengue Fever disease. To help the Government of West Java, this research conducted a grouping of dengue fever in West Java in 2016-2021. This research uses the Knowledge Discovery in Database (KDD) method. The algorithm used is k-means clustering with the help of the elbow method to get the optimal number of clusters, which is 2 clusters. Cluster 0 with low category consists of 22 regions, and cluster 2 with high category consists of 5 regions. The result of silhouette coefficient evaluation is 0.689 with standard structure criteria.

Keywords: *Elbow method, K-means Clustering, Knowledge Data Discovery.*

Correspondence: dea.amelia18151@student.unsika.ac.id

How to Cite: Amelia, D., Padilah, T., & Jamaludin, A. (2022). Optimasi Algoritma K-Means Menggunakan Metode Elbow dalam Pengelompokan Penyakit Demam Berdarah Dengue (DBD) di Jawa Barat. *Jurnal Ilmiah Wahana Pendidikan*, 8(11), 207-215. <https://doi.org/10.5281/zenodo.6831380>

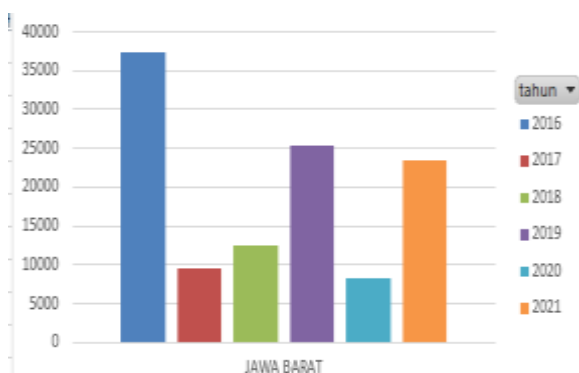
PENDAHULUAN

Demam Berdarah *Dengue* (DBD) merupakan penyakit menular demam akut yang biasa terjadi di daerah tropis dan subtropis di dunia dengan disebabkan oleh virus yang ditularkan oleh nyamuk *Aedes*, yaitu *Aedes Aegypti* dan *Aedes Albopictus* (Widyatami & Suryawan, 2021). Faktor terjadinya DBD sebagian besar terjadi karena faktor lingkungan, dimana nyamuk *Aedes Aegypti* senang berada pada lingkungan dengan kualitas buruk, misalnya lingkungan yang lembab, pemukiman yang penduduknya padat serta kurang cahaya matahari, banyak tumpukan sampah, gelap, terdapat genangan air sehingga menjadi tempat berkembang biaknya nyamuk, seperti drum air, bak mandi, maupun barang bekas yang sudah tidak terpakai, dan sebagainya (Kiki Fatmawati, 2018).

Penyakit demam berdarah ini gejala nya sangat luas, sehingga banyak masyarakat yang sulit mendeteksi gejala awal dari demam berdarah. Gejala demam berdarah yang paling sering ditemukan yaitu terjadinya pendarahan atau ruam pada kulit (Sembiring, 2022). Dampak yang terjadi ketika demam berdarah yaitu penderita mengalami kenaikan suhu tubuh yang sangat tinggi dengan disertai demam, sakit kepala, mual dan muntah, serta nyeri pada bagian perut (Hariyanto & Shita, 2018). Bahkan, jika penyakit demam berdarah ini telat dalam penanganan dapat menyebabkan kehilangan nyawa bagi penderitanya.

Menurut *World Health Organization* (WHO), kasus DBD diperkirakan setiap tahunnya mencapai 50 sampai dengan 100 juta terinfeksi dan total kematian akibat DBD dapat mencapai 22.0000 jiwa. Penyakit DBD juga menjadi salah satu penyakit endemik di lebih dari 100 negara di seluruh dunia, seperti di Afrika, Mediterania Timur, Pasifik Barat, Amerika, beberapa Negara di Eropa, dan Asia Tenggara. Benua Asia menjadi benua yang memiliki kasus demam berdarah paling tertinggi di dunia setiap tahunnya. Di Asia Tenggara sendiri, penyakit demam berdarah mencapai 1,3 miliar atau sebanding dengan 52% dari 2,5 miliar jiwa di seluruh dunia yang memiliki resiko penyakit DBD. Sementara itu, sejak tahun 1968 sampai tahun 2011, WHO mencatat bahwa Indonesia menjadi Negara dengan kasus demam berdarah paling tinggi di kawasan Asia Tenggara (Akbar & Masulana Syaputra, 2019).

Menurut Kementerian Kesehatan Indonesia, tercatat pada tahun 2022 terhitung sampai bulan Juni kasus DBD mencapai 45.387 kasus, dengan total kasus kematian hingga 432 kasus. Salah satu provinsi dengan kasus demam berdarah terbanyak yaitu Jawa Barat. Menurut Open Data Jawa Barat, kasus demam berdarah di Jawa Barat mengalami naik turun dari tahun 2016-2021. Pada tahun 2020 kasus demam berdarah sempat mengalami penurunan, tetapi kasus meningkat kembali pada tahun 2021 hingga 4 kali lipat. Grafik kasus demam berdarah disajikan pada Gambar 1.



Gambar 1. Grafik Penyakit DBD Jawa Barat Tahun 2016-2021

Gambar 1 menunjukkan grafik penyakit DBD di Jawa Barat dari tahun 2016-2021, dimana tahun 2016 merupakan tahun tertinggi kasus DBD dengan total 37.279 kasus. Pada tahun 2017-2018 mengalami penurunan yang cukup drastis. Tetapi, pada tahun 2019 kasus demam berdarah kembali meningkat hingga mencapai 25.282 kasus. Pada tahun 2020 kasus DBD kembali menurun. Pada tahun 2021 kasus demam berdarah kembali mengalami peningkatan yang cukup drastis hingga mencapai 23.454 kasus. Dengan kasus yang cukup tinggi ini, Pemerintah Jawa Barat harus memberikan perhatian khusus pada penyakit DBD karena jika tidak ditangani dengan serius akan menyebabkan kematian bagi penderitanya.

Dalam upaya membantu Pemerintah Jawa Barat khususnya Dinas Kesehatan Provinsi Jawa Barat, penelitian ini akan melakukan pengelompokan penyakit DBD di Jawa Barat. Dengan dilakukannya analisis pengelompokan penyakit DBD ini, diharapkan hasil penelitian dapat menjadi dasar dalam pengambilan keputusan dalam penanganan penyakit DBD secara cepat dan efektif. Pengelompokan penyakit DBD dapat dilakukan dengan menggunakan salah satu teknik *data mining* yaitu algoritma *k-means clustering*.

K-means clustering merupakan salah satu algoritma *clustering* non hirarki dengan melakukan pembagian data ke dalam beberapa *cluster* sehingga data akan dikelompokkan

pada *cluster* yang sama jika memiliki karakteristik yang hampir sama, sedangkan data dengan karakteristik berbeda akan dikelompokkan pada *cluster* lain (Aulia, 2021). Dengan menggunakan metode *k-means clustering*, diharapkan pihak Dinas Kesehatan Jawa Barat dapat mengetahui pengelompokan penyakit DBD dan mengetahui Kabupaten/Kota mana saja yang memiliki angka kasus DBD tinggi, sehingga kedepannya dapat memprioritaskan penanganan penyakit pada Kabupaten/Kota dengan kasus DBD tinggi tersebut.

Pada algoritma *k-means*, dalam menentukan jumlah *cluster* dapat dilakukan secara acak, tetapi hal tersebut dapat menurunkan hasil evaluasi karena memungkinkan kesalahan dalam pemilihan jumlah *cluster*. Untuk meminimalisir hal tersebut terjadi, pemilihan *cluster* dapat menggunakan metode *elbow*. Metode *elbow* merupakan metode yang digunakan untuk menentukan jumlah *cluster* terbaik yang berdasarkan pembentukan sudut siku di satu titik dari hasil persentase perbandingan antara jumlah klaster (k).

Banyak penelitian sebelumnya yang membahas tentang pengelompokan penyakit dengan menggunakan algoritma *k-means clustering*, diantaranya yaitu penelitian oleh (Sari et al., 2019) yang melakukan pengelompokan penyakit ISPA di Kota Sekayu dengan menggunakan data tahun 2017-2018. Algoritma *k-means* berhasil membagi data menjadi 3 *cluster* yaitu *cluster* 0 dengan kategori rendah berjumlah 111 data, *cluster* 1 dengan kategori sedang berjumlah 243 data, dan *cluster* 2 dengan kategori tinggi dengan total 791 data.

Penelitian selanjutnya oleh Yanti Puspita Sari, Aji Primajaya, & Agung Susilo Yuda Irawan (2020) dengan melakukan pengelompokan penyakit *tuberculosis* di Kabupaten Karawang tahun 2018. Pemilihan *cluster* menggunakan metode *elbow* dengan menghasilkan 3 *cluster*. Algoritma yang digunakan yaitu *k-means clustering* dimana hasil pemodelan terbagi menjadi 3 *cluster*, dimana *cluster* 0 memiliki kategori tinggi, *cluster* 1 dengan kategori sedang, dan *cluster* 2 dengan kategori rendah. Hasil evaluasi pemodelan *k-means* dilakukan dengan dua metode yaitu SSE dan *silhouette coefficient*. Pada evaluasi SSE dengan jumlah $k=3$ dan *random seed*=10 menghasilkan nilai evaluasi 2,4402. Sedangkan untuk evaluasi *silhouette coefficient* menghasilkan nilai sebesar 0,5629 yang masuk ke dalam kategori standar.

Berdasarkan latar belakang tersebut, penelitian ini akan melakukan pengelompokan penyakit DBD di Provinsi Jawa Barat dengan menggunakan algoritma *k-means clustering* dan juga metode *elbow* untuk mendapatkan jumlah *cluster* optimal dan sebagai upaya meningkatkan hasil evaluasi model.

METODOLOGI PENELITIAN

Pada penelitian ini menggunakan metodologi *Knowledge Discovery in Database* (KDD) dengan enam tahapan yaitu *data selection*, *data preprocessing*, *data transformation*, *data mining*, dan *evaluation*. Rincian metodologi KDD sebagai berikut:

1. *Data Selection*

Tahap pertama yaitu *data selection* dimana melakukan pemilihan data serta pemilihan atribut yang akan digunakan pada proses pemodelan nanti. Tidak semua atribut pada data digunakan karena hanya data yang relevan serta sesuai dengan kebutuhan dan tujuan penelitian saja yang akan digunakan.

2. *Data Preprocessing*

Tahap *data preprocessing* merupakan tahap dimana data dilakukan pembersihan atau biasa disebut dengan *data cleaning*. *Data preprocessing* dilakukan dengan

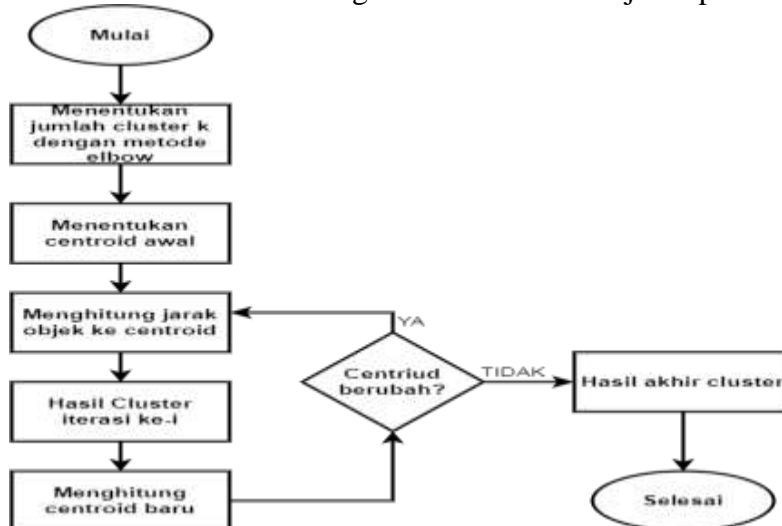
menghilangkan *missing value*, data inkonsisten, maupun *noise data*. Dengan adanya data *preprocessing* ini berupaya untuk meningkatkan hasil evaluasi.

3. *Data Transformation*

Pada tahap *data transformation* atau transformasi data ini data dilakukan proses pengubahan format data ke bentuk format dataset baru menyesuaikan dengan kebutuhan dan tujuan penelitian. Proses transformasi ini berguna untuk memudahkan proses pemodelan nantinya.

4. *Data Mining*

Tahap *data mining* ini menerapkan teknik pemodelan *data mining* yang akan digunakan pada penelitian. Pada penelitian ini menggunakan algoritma *k-means clustering*. Dalam menentukan jumlah *cluster* optimal, penelitian ini menggunakan metode *elbow*. Alur dari algoritma *k-means* disajikan pada Gambar 2.



Gambar 2. Alur Algoritma *K-means*

5. *Evaluation*

Setelah tahap *data mining*, tahap selanjutnya yaitu *evaluation* atau evaluasi terhadap hasil pemodelan yang telah dibentuk. Pada tahap evaluasi ini menganalisis hasil model *k-means clustering*. Penelitian ini menggunakan metode evaluasi *silhouette coefficient* yang menguji kualitas dari *cluster* yang telah dihasilkan sehingga dapat mengetahui apakah model sudah sesuai dengan tujuan penelitian yang ingin dicapai.

Alur dari metodologi KDD dapat dilihat pada Gambar 3.



Gambar 3. Alur Metodologi KDD

HASIL DAN PEMBAHASAN

Data Selection

Tahap pertama yaitu *data selection* dengan melakukan pemilihan atribut pada *dataset*. Data yang digunakan pada penelitian ini yaitu data penyakit demam berdarah *dengue* berdasarkan kabupaten/kota di Jawa Barat tahun 2016-2021. *Dataset* didapatkan dari situs resmi Open Data Jabar. *Dataset* awal terdiri dari 8 atribut yaitu *id*, *kode_provinsi*, *nama_provinsi*, *kode_kabupaten_kota*, *nama_kabupaten_kota*, *jumlah_kasus*, *satuan*, *tahun*. Atribut yang dipilih hanya 3 atribut yaitu *nama_kabupaten_kota*, *jumlah_kasus*, *tahun*. Atribut yang dipilih hanya yang relevan dengan tujuan penelitian. *Dataset* yang telah melalui *data selection* dapat dilihat pada Gambar 4.

nama_kabupaten_kota	jumlah_kasus	tahun
KABUPATEN BOGOR	3477	2016
KABUPATEN SUKABUMI	854	2016
KABUPATEN CIANJUR	799	2016
KABUPATEN BANDUNG	3466	2016
KABUPATEN GARUT	561	2016
KABUPATEN TASIKMALAYA	478	2016
KABUPATEN CIAMIS	748	2016
KABUPATEN KUNINGAN	1720	2016
KABUPATEN CIREBON	1877	2016
KABUPATEN MAJALENGKA	322	2016
KABUPATEN SUMEDANG	944	2016
KABUPATEN INDRAMAYU	911	2016
KABUPATEN SUBANG	699	2016
KABUPATEN PURWAKARTA	772	2016
KABUPATEN KARAWANG	1059	2016
KABUPATEN BEKASI	1802	2016
KABUPATEN BANDUNG BARU	1545	2016
KABUPATEN PANGANDARAI	137	2016
KOTA BOGOR	1229	2016
KOTA SUKABUMI	844	2016

Gambar 4. Hasil *Data Selection*

Data Preprocessing

Tahap selanjutnya yaitu *data preprocessing* dimana data dilakukan pembersihan dari *missing value*. Untuk *dataset* sendiri memiliki *missing value* ketika dilakukan pengecekan dengan *library python* yang dapat dilihat pada Gambar 5.

```
nama_kabupaten_kota    0
jumlah_kasus          9
tahun                  0
.....
```

Gambar 5. Pengecekan *Missing Value*

Dapat dilihat pada Gambar 5, atribut *jumlah_kasus* memiliki 9 *missing value*, sementara 2 atribut lainnya tidak terdapat *missing value*. *Missing value* pada atribut *jumlah_kasus* harus ditangani agar tidak mempengaruhi hasil pemodelan. *Missing value* ditangani dengan menggantikan nilai yang kosong dengan nilai *mean* (rata-rata). Setelah dilakukan proses *preprocessing*, *dataset* tidak memiliki *missing value* seperti pada Gambar 6.

```
nama_kabupaten_kota    0
jumlah_kasus          0
tahun                  0
.....
```

Gambar 6. Hasil Penanganan *Missing Value*

Dapat dilihat pada Gambar 6, *dataset* tidak lagi terdapat *missing value*. Untuk itu data sudah siap digunakan untuk melalui tahap selanjutnya.

Data Transformation

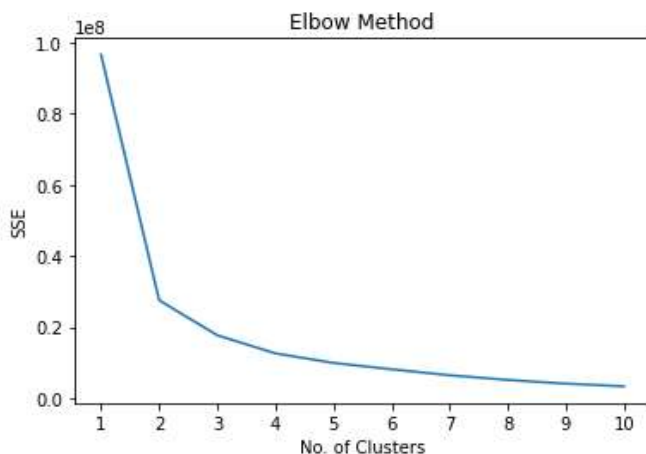
Pada tahap *data transformation* ini, *dataset* dilakukan transformasi data dengan membuat format *dataset* baru dari *dataset* yang telah dipilih sebelumnya. *Dataset*

sebelumnya masih memiliki format atribut yang kurang mendukung tujuan penelitian, karena atribut tahun hanya berisi nilai nominal, bukan nilai numerik yang dapat dihitung. Untuk itu atribut tahun perlu ditransformasikan ke dalam bentuk yang lebih sesuai dan mendukung penelitian. Atribut tahun dipecah dan dipisahkan setiap tahunnya dan digabungkan dengan atribut jumlah kasus, sehingga menghasilkan atribut baru yaitu total kasus demam berdarah berdasarkan tahun. Dengan dilakukannya transformasi data ini dapat memudahkan proses pemodelan. *Dataset* hasil transformasi dapat dilihat pada Gambar 7

nama_kabupaten_kota	tahun 2016	tahun 2017	tahun 2018	tahun 2019	tahun 2020	tahun 2021
KABUPATEN BOGOR	3477	277	741	1210	760,27451	2220
KABUPATEN SUKABUMI	854	256	233	616	561	272
KABUPATEN CIANJUR	799	0	113	643	684	430
KABUPATEN BANDUNG	3466	1013	1774	2635	760,27451	2026
KABUPATEN GARUT	561	406	331	769	651	1011
KABUPATEN TASIKMALAYA	478	224	196	210	179	48
KABUPATEN CIAMIS	748	47	29	404	760,27451	470
KABUPATEN KUNINGAN	1720	728	352	505	362	544
KABUPATEN CIREBON	1877	274	215	1291	880	819
KABUPATEN MAJALENGA	322	112	108	457	760,27451	447
KABUPATEN SUMEDANG	944	184	408	663	707	1264
KABUPATEN INDRAMAYU	911	113	911	222	214	188
KABUPATEN SUBANG	699	104	143	270	283	287
KABUPATEN PURWAKARTA	772	122	263	544	534	220
KABUPATEN KARAWANG	1059	102	60	199	760,27451	929
KABUPATEN BEKASI	1802	332	181	511	317	370
KABUPATEN BANDUNG BARAT	1545	201	419	1706	968	419
KABUPATEN PANGANDARAN	137	175	75	158	196	79
KOTA BOGOR	1229	855	727	611	241	526
KOTA SUKABUMI	844	352	238	786	651	449

Gambar 7. Hasil Transformasi Data *Data Mining*

Setelah data sudah siap digunakan, tahap selanjutnya yaitu implementasi teknik *data mining*. Teknik *data mining* yang digunakan pada penelitian ini yaitu algoritma *k-means*. Dalam algoritma *k-means*, harus ditentukan jumlah *cluster*-nya, salah satunya dengan menggunakan metode *elbow*. Dengan menggunakan metode *elbow* dapat diketahui hasil *cluster* optimal yang dapat dilihat melalui titik sudut siku pada grafik *elbow*. Hasil grafik metode *elbow* dapat dilihat pada Gambar 8.



Gambar 8. Hasil Grafik *Elbow*

Berdasarkan Gambar 8, grafik metode *elbow* menunjukkan bahwa jumlah *cluster* yang paling optimal adalah 2 karena pada titik 2 grafik lebih membentuk sudut siku, sehingga menjadi *cluster* yang paling optimal dibandingkan dengan nilai lainnya.

Setelah mengetahui hasil *cluster* optimal ($k=2$), selanjutnya membentuk model *k-means* dengan *library python*. Jika titik *centroid* konvergen atau tidak berubah dari nilai sebelumnya, maka nilai tersebut menjadi titik *centroid* akhir. *Centroid* akhir dari *k-means* dapat dilihat pada Tabel 1.

Tabel 1. Centroid Akhir K-means

Cluster	tahun 2016	tahun 2017	tahun 2018	tahun 2019	tahun 2020	tahun 2021
C0	900,682	241,545	256,091	565,455	509,959	475,727
C1	3492,8	864,6	1371,6	2568,4	760,275	2597,6

Setelah mengetahui hasil *centroid* akhir *k-means*, maka pemodelan sudah terbentuk dan data sudah masuk ke dalam *cluster*. Ada dua kategori *cluster* yaitu *cluster* 0 dengan kategori rendah dan *cluster* 1 dengan kategori tinggi. Hasil pengelompokan *k-means* dapat dilihat pada Tabel 2.

Tabel 2. Hasil Pengelompokan K-means

Cluster	Nama Daerah	Kategori
0	Kab. Sukabumi, Kab. Cianjur, Kab. Garut, Kab. Tasikmalaya, Kab. Ciamis, Kab. Kuningan, Kab. Cirebon, Kab. Majalengka, Kab. Sumedang, Kab. Indramayu, Kab. Subang, Kab. Purawakarta, Kab. Karawang, Kab. Bekasi, Kab. Bandung Barat, Kab. Pangandaran, Kota Bogor, Kota Sukabumi, Kota Cirebon, Kota Cimahi, Kota Tasikmalaya, Kota Banjar	Rendah
1	Kota Depok, Kota Bekasi, Kota Bandung, Kab. Bandung, Kab. Bogor	Tinggi

Berdasarkan Tabel 2, dapat diketahui bahwa hasil *clustering k-means* menghasilkan 2 *cluster* dengan dua kategori yaitu *cluster* 0 dengan kategori rendah yang terdiri dari 22 daerah kabupaten/kota, dan *cluster* 1 dengan kategori tinggi yang terdiri dari 5 daerah kabupaten/kota. Hasil *k-means* divisualisasikan ke dalam bentuk grafik seperti pada Gambar 9.



Gambar 9. Grafik Hasil K-means DBD Jawa Barat tahun 2016-2021

Hasil pengelompokan ini dapat dijadikan sebagai dasar pengambilan keputusan Pemerintahan Jawa Barat dalam penanganan kasus DBD ini, khususnya diprioritaskan bagi daerah yang kategori kasusnya tinggi sehingga menekan angka kasus DBD di daerah tersebut.

Evaluation

Setelah didapatkan hasil *k-means clustering*, selanjutnya yaitu tahap evaluasi dengan menggunakan metode *silhouette coefficient*. Dengan menggunakan *silhouette coefficient* ini dapat melihat kualitas dari *cluster* yang telah terbentuk. Untuk dapat memudahkan menganalisis hasil evaluasi, *silhouette coefficient* memiliki kriteria penilaian yang dapat dilihat pada Tabel 3 berikut.

Tabel 3. Kriteria Nilai *Silhouette Coefficient*

Nilai <i>Silhouette coefficient</i>	Kriteria Penilaian
$0,7 < SC \leq 1,0$	Struktur kuat
$0,5 < SC \leq 0,7$	Struktur standard
$0,25 < SC \leq 0,5$	Struktur lemah
$SC \leq 0,25$	Tidak memiliki struktur

Proses evaluasi *silhouette coefficient* pada hasil *k-means clustering* dapat dilihat pada Gambar 10.

Nilai *Silhouette Coefficient*: 0.6892567683426529

Gambar 10. Hasil *Silhouette Coefficient*

Berdasarkan Gambar 10 hasil *silhouette coefficient* sebesar 0.689, maka hasil evaluasi masuk ke dalam kategori struktur standar. Dengan kata lain, setiap data yang masuk pada masing-masing cluster sudah masuk pada cluster yang tepat dan hasil algoritma *k-means* sudah baik.

KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, maka dapat disimpulkan sebagai berikut:

1. Algoritma *k-means clustering* dengan menggunakan metode *elbow* mampu menghasilkan 2 *cluster* optimal dalam pengelompokan penyakit Demam Berdarah Dengue (DBD) di Jawa Barat tahun 2016-2021.
2. Hasil *clustering* terbagi menjadi 2 *cluster* yaitu *cluster* 0 dengan kategori rendah yang terdiri dari 22 daerah kabupaten/kota, dan *cluster* 1 dengan kategori tinggi yang terdiri dari 5 daerah kabupaten/kota.
3. Evaluasi *cluster* dengan menggunakan *silhouette coefficient* menghasilkan nilai sebesar 0.689 yang artinya *cluster* masuk ke dalam kategori standar dan data sudah masuk ke dalam *cluster* yang tepat.

DAFTAR PUSTAKA

Akbar, H., & Maulana Syaputra, E. (2019). Faktor Risiko Kejadian Demam Berdarah Dengue (DBD) di Kabupaten Indramayu. *MPPKI (Media Publikasi Promosi Kesehatan Indonesia): The Indonesian Journal of Health Promotion*, 2(3), 159–164. <https://doi.org/10.31934/mppki.v2i3.626>

Hariyanto, M., & Shita, R. T. (2018). Clustering pada Data Mining untuk Mengetahui Potensi Penyebaran Penyakit DBD Menggunakan Metode Algoritma K-Means dan Metode Perhitungan Jarak Euclidean Distance. *Sistem Komputer Dan Teknik*

Informatika, 1(1), 117–122.

- Kiki Fatmawati, A. P. W. (2018). Data Mining : Penerapan Rapidminer Dengan K-Means Cluster Pada Daerah Terjangkit Demam Berdarah Dengue (Dbd) Berdasarkan Provinsi. *CESS (Journal of Computer Engineering System and Science*, 3(2), 173–178.
- Sari, T. A., Megawaty, M., & Putra, A. (2019). Pengelompokan Penyebaran Penyakit Ispa Di Wilayah Kota Sekayu Menggunakan Algoritma K-Means Clustering (Studi Kasus: Rsud Sekayu). *Bina Darma Conference on Computer Science (BDCCS)*, 1(1), 174–184.
- Sembiring, E. M. (2022). Penerapan K-Means Clustering Untuk Pengelompokan Penyebaran Demam Berdarah Dengue (DBD) Di Kabupaten Deli Serdang. *TIN: Terapan Informatika Nusantara*, 2(11), 673–677. <https://doi.org/10.47065/tin.v2i11.1503>
- Widyatami, A. I., & Suryawan, D. A. (2021). Pengelompokan Daerah Rawan Demam Berdarah Dengue di Provinsi DKI Jakarta. *Indonesian of Health Information Management Journal (INOHIM)*, 9(1), 73–82. <https://doi.org/10.47007/inohim.v9i1.241>
- Kasus DBD Meningkat, Kemenkes Galakkan Gerakan 1 Rumah 1 Jumantik (G1R1J). (15 Juni 2022). Sehat Negeriku Kemenkes RI. <https://sehatnegeriku.kemkes.go.id/baca/umum/20220615/0240172/kasus-dbd-meningkat-kemenkes-galakkan-gerakan-1-rumah-1-jumantik-g1r1j/#:~:text=Kementerian%20Kesehatan%20mencatat%20di%20tahun,akibat%20DBD%20mencapai%20432%20kasus.>
- Jumlah Kasus Penyakit Demam Berdarah Dengue (DBD) Berdasarkan Kabupaten/Kota di Jawa Barat. Open Data Jabar. Diakses pada 30 Juni 2022. <https://opendata.jabarprov.go.id/id/dataset/jumlah-kasus-penyakit-demam-berdarah-dengue-dbd-berdasarkan-kabupatenkota-di-jawa-barat>