

Analisis Daerah Potensi Persebaran Demam Berdarah Dengue di DKI Jakarta Menggunakan K-Means

Hani Dzikra Nurkhairiyah*¹, Ati Zaidiah², Helena Nurramdhani Irmanda³
Program Studi S1 Sistem Informasi, Fakultas Ilmu Komputer, UPN Veteran Jakarta
e-mail: ¹hanidzikra21@gmail.com, ²atizaidiah@upnvj.ac.id, ³helenairmanda@upnvj.ac.id
*Penulis Korespondensi

Diterima: 28 Oktober 2022; Direvisi: 8 Juli 2024; Disetujui: 8 Juli 2024

Abstrak

Salah satu penyakit yang masih belum bisa dibendung di Indonesia adalah demam berdarah dengue, penyakit ini dibawa oleh gigitan nyamuk *Aedes Aegypti* dan *Aedes Albopictus*. Motivasi di balik penelitian ini adalah untuk mengkarakterisasi daerah yang diperkirakan menjadi titik penyebaran demam berdarah, khususnya di wilayah DKI Jakarta. Hal ini perlu disadari oleh pemerintah maupun masyarakat setempat mengenai pentingnya peningkatan kewaspadaan terhadap penyakit demam berdarah. Teknik data mining digunakan pada penelitian ini, serta clustering sebagai metodenya dengan menerapkan algoritma k-means. Nilai K yang digunakan pada penelitian ini sudah dievaluasi oleh Silhouette yang menghasilkan Silhouette Score sebesar 0,41642 pada K=2, maka cluster pada penelitian ini dibagi menjadi 2 cluster. Clustering menghasilkan daerah sporadis dan daerah endemis. Daerah sporadis mencakup 30 kecamatan dan daerah endemis mencakup 14 kecamatan.

Kata kunci: data mining, clustering, k-means, demam berdarah dengue

Abstract

An irresistible illness that is as yet present in Indonesia is dengue hemorrhagic fever, this sickness is brought about by the chomp of the *Aedes Aegypti* and *Aedes Albopictus* mosquitoes. The motivation behind this study is to characterize areas of expected spread of dengue fever, particularly in the DKI Jakarta region. The government and local communities need to realize the importance of increasing awareness of dengue fever. Data mining is used in this study, as well as clustering as a method by applying the k-means algorithm. The K value used in this study has been evaluated by Silhouette which produces a Silhouette Score of 0.41642 at K = 2, then the cluster in this study is divided into 2 clusters. Clustering produces sporadic areas and endemic areas. Sporadic areas include 30 sub-districts and endemic areas include 14 sub-districts.

Keywords: data mining, clustering, k-means, dengue hemorrhagic fever.

1. PENDAHULUAN

Demam Berdarah Dengue (DBD) merupakan salah satu penyakit tak tertahankan yang ada di Indonesia. Demam berdarah disebabkan oleh salah satu infeksi *dengue* yang berasal dari gigitan nyamuk *Aedes Aegypti* dan *Aedes albopictus*. Penyebaran demam berdarah di wilayah DKI Jakarta dipengaruhi oleh curah hujan serta kelembaban [1]. Curah hujan tinggi dapat mengakibatkan banjir yang secara tidak langsung berpengaruh terhadap perubahan iklim [2],

maka dari itu banjir menjadi salah satu variabel yang menjadi penyebab penyebaran demam berdarah, karena banjir mempengaruhi penyebaran demam berdarah secara tidak langsung [3].

Variabel lainnya yang berpengaruh terhadap penyebaran demam berdarah ialah kepadatan penduduk. Penyebaran demam berdarah lebih rentan terjadi di daerah dengan kepadatan penduduk yang lebih tinggi. Mobilitas penduduk mengakibatkan persebaran demam berdarah menjadi meningkat dimulai dari satu titik kemudian ke titik berikutnya, dengan perkiraan jarak terbang nyamuk sekitar 50 meter [4]. DKI Jakarta menjadi salah satu daerah dengan kepadatan penduduk yang tinggi, disertai dengan aktivitas penduduk yang padat.

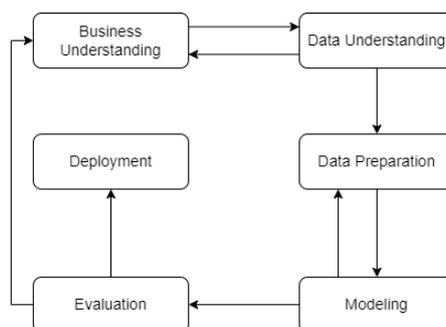
Kasus demam berdarah *dengue* di Indonesia semakin bertambah, sudah tercatat sebanyak 71.633 kasus terhitung hingga Juli 2020. Kasus demam berdarah terbanyak terletak pada 10 provinsi. Pada provinsi-provinsi tersebut, dari tahun ke tahun potensi menjadi daerah endemis semakin tinggi. Salah satu daerah dengan kasus demam berdarah tertinggi ialah daerah DKI Jakarta sebanyak 4.227 kasus [5], hal ini perlu disadari oleh pemerintah maupun masyarakat setempat mengenai pentingnya peningkatan kewaspadaan terhadap penyakit demam berdarah. Melihat tingginya angka kasus penyebaran demam berdarah di DKI Jakarta, Ketua Komisi E DPRD Iman Satria mengatakan bahwa sebaiknya Dinas Kesehatan waspada terhadap penyebaran demam berdarah di DKI Jakarta [6].

Melihat permasalahan tersebut, penelitian ini dilakukan guna mengelompokkan daerah potensi persebaran demam berdarah di DKI Jakarta, sehingga pihak pemerintah serta masyarakat mengetahui daerah mana saja yang termasuk ke dalam daerah endemis maupun daerah sporadis. Hal tersebut diharapkan dapat membantu pihak Dinas Kesehatan DKI Jakarta untuk meningkatkan kewaspadaan serta dapat menekan angka DBD dengan melakukan upaya-upaya pencegahan DBD sebelum angka persebaran DBD di daerah DKI Jakarta semakin melonjak naik. *Data mining* yang digunakan pada penelitian ini guna mengelompokkan daerah potensi persebaran demam berdarah ialah *clustering*, serta menerapkan algoritma *K-Means*. Terdapat beberapa keunggulan *K-Means*, seperti ide matematika yang sederhana, serta implementasi yang mudah [7].

Penerapan algoritma *K-Means* sudah berhasil diterapkan pada beberapa penelitian, di antaranya pada penelitian yang berjudul “Penerapan Algoritma *K-means Clustering* untuk Pengelompokan Penyebaran *Covid-19* di Provinsi Jawa Barat”. Penelitian tersebut menghasilkan *cluster* penyebaran *Covid-19*, di antaranya C1 dengan angka penyebaran *Covid-19* tertinggi di daerah Depok dan Bekasi, C2 dengan angka penyebaran *Covid-19* sedang yang berada di daerah Kota Bandung, Kabupaten Bandung, Kabupaten Bekasi, Kabupaten Bogor, dan Kabupaten Karawang, serta C3 dengan angka *Covid-19* yang lebih sedikit [8].

2. METODE PENELITIAN

Teknik yang digunakan dalam penelitian ini adalah CRISP-DM. CRISP-DM (*Cross-Industry Standard Process for Data Mining*) adalah metodologi penambangan informasi yang terdiri dari beberapa fase pemeriksaan untuk menangani masalah.



Gambar 1. Tahapan CRISP-DM

Tahapan dalam CRISP-DM terdiri dari pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan penyebaran [9]. Tahapan CRISP-DM yang diimplementasikan pada penelitian ini terlihat lebih jelas pada gambar 1.

2.1 Pemahaman Bisnis (*Business Understanding*)

Tahapan pertama yang dilakukan pada penelitian ini adalah mengidentifikasi masalah. Proses yang dilakukan ialah dengan cara membaca beberapa artikel jurnal, sehingga menemukan suatu fenomena yang dapat dipecahkan dengan menggunakan *data mining*.

2.2 Pemahaman Data (*Data Understanding*)

Kegiatan yang dilakukan pada tahapan ini adalah pengumpulan data yang akan digunakan untuk penelitian serta cara pemahaman data yang sudah didapatkan. Data yang didapatkan dan akan digunakan ialah data penderita demam berdarah di DKI Jakarta, data kepadatan penduduk DKI Jakarta, serta data daerah rawan banjir di DKI Jakarta.

2.3 Persiapan Data (*Data Preparation*)

Kegiatan yang diterapkan dalam tahapan ini ialah penentuan *feature* yang akan dipakai, serta melakukan *normalisasi min-max* dengan tujuan untuk mengubah data menjadi nilai dengan *range* 0-1 dengan menggunakan *Python*.

2.4 Pemodelan (*Modeling*)

Penelitian ini menggunakan algoritma *k-means* guna melakukan pengelompokan daerah potensi persebaran demam berdarah *dengue* di DKI Jakarta. Adapun cara kerja dari algoritma *k-means* yaitu [9]: 1) Menentukan nilai K yang akan dibagi; 2) Menetapkan nilai *centroid* awal secara acak; 3) Menghitung jarak minimum tiap data terhadap *centroid*; 4) Menghitung nilai *centroid* yang baru; 5) Mengulangi langkah 3 hingga data tidak ada yang berpindah *cluster*.

2.5 Evaluasi (*Evaluation*)

Hasil *clustering* yang didapatkan, dievaluasi menggunakan *Silhouette Coeficient*. *Silhouette Coeficient* digunakan untuk mengetahui kualitas dari hasil *clustering*. Pembagian *cluster* atau nilai K yang baik ditunjukkan oleh nilai *Silhouette Score* yang mendekati 1 [10].

2.6 Penyebaran (*Deployment*)

Pada tahapan ini, hasil cluster yang didapatkan divisualisasikan dalam bentuk *maps* daerah potensi persebaran demam berdarah di DKI Jakarta.

3. HASIL DAN PEMBAHASAN

3.1. Pemahaman Bisnis (*Business Understanding*)

Proses awal yang dilakukan adalah mengidentifikasi masalah dengan cara membaca beberapa artikel jurnal, selain itu juga menghimpun beberapa literatur yang dibutuhkan guna mendapatkan informasi baru yang diperlukan sehingga proses selanjutnya ialah menyusun rumusan permasalahan untuk penelitian. Motivasi di balik penelitian ini adalah untuk mengkarakterisasi daerah potensi persebaran DBD di DKI Jakarta.. Hasil dari penelitian ini diharapkan dapat bermanfaat bagi pihak Pemerintah DKI Jakarta guna mengambil keputusan dalam upaya menekan angka penyebaran DBD sebelum angka kasus melonjak naik.

3.2. Pemahaman Data (*Data Understanding*)

Data yang digunakan pada penelitian ini merupakan data sekunder yang diambil melalui situs Open Data Jakarta melalui laman URL <https://data.jakarta.go.id/>. Situs tersebut disediakan oleh pemerintah DKI Jakarta berisi data-data DKI Jakarta. Berdasarkan latar belakang yang menjelaskan mengenai faktor-faktor dari penyebaran demam berdarah, bahwa banjir merupakan salah satu faktor secara tidak langsung terhadap penyebaran demam berdarah [3], serta faktor lainnya ialah kepadatan penduduk [4]. Berdasarkan hal tersebut, maka pada penelitian ini sumber data yang digunakan ialah data penderita demam berdarah di DKI Jakarta, data kepadatan penduduk DKI Jakarta, serta data daerah rawan banjir di DKI Jakarta. Seluruh data yang digunakan merupakan data pada tahun 2020. Variabel yang akan digunakan pada penelitian ini dapat dilihat pada tabel 1 berikut.

Tabel 1. Variabel yang digunakan

| Nama Variable | Keterangan | Sumber Data |
|---------------------|--|---|
| Kota_administrasi | Kota di wilayah DKI Jakarta | Dinas Kesehatan DKI Jakarta |
| Kecamatan | Kecamatan di wilayah DKI Jakarta | Dinas Kesehatan DKI Jakarta |
| Penderita_dbd | Banyaknya penderita DBD di DKI Jakarta berdasarkan kecamatan | Dinas Kesehatan DKI Jakarta |
| Luas_wilayah | Jumlah luas wilayah DKI Jakarta per kecamatan (kilometer persegi) | Dinas Kependudukan dan Pencatatan Sipil |
| Kepadatan_penduduk | Jumlah kepadatan penduduk DKI Jakarta Berdasarkan jumlah penduduk/luas wilayah | Dinas Kependudukan dan Pencatatan Sipil |
| Rw_terdampak_banjir | Jumlah RW yang terdampak banjir | BPBD Provinsi DKI Jakarta |
| Rt_terdampak_banjir | Jumlah RT yang terdampak banjir | BPBD Provinsi DKI Jakarta |

3.3. Persiapan Data (*Data Preparation*)

Kegiatan yang dikerjakan pada proses ini adalah seleksi *feature* yang akan digunakan serta melakukan normalisasi *min-max*. Pada data yang ada, terdapat *range* nilai yang jauh antar variabel, sehingga perlu dilakukannya normalisasi *min-max*. Normalisasi *min-max* ini dilakukan dengan menggunakan perintah *python*. Berikut rincian *dataset* yang digunakan setelah melalui proses seleksi *feature*, dapat dilihat pada tabel 2.

Tabel 2. *Dataset* Setelah diolah

| kecamatan | penderita_dbd | luas_wilayah | kepadatan_penduduk | rw_terdampak_banjir | rt_terdampak_banjir |
|--------------------------|---------------|--------------|--------------------|---------------------|---------------------|
| Cempaka Putih | 51 | 4,70 | 72335,41 | 0,75 | 0,83 |
| Gambir | 19 | 7,59 | 105915,63 | 0,08 | 0,08 |
| Johar Baru | 39 | 2,37 | 297047,82 | 0,75 | 0,83 |
| Kemayoran | 92 | 7,25 | 339115,19 | 2,17 | 2,58 |
| Menteng | 38 | 6,53 | 77598,95 | 0,50 | 0,50 |
| ... | ... | ... | ... | ... | ... |
| Pasar Rebo | 146 | 12,97 | 90465,78 | 1,83 | 2,00 |
| Pulo Gadung | 113 | 15,60 | 155433,61 | 3,50 | 7,67 |
| Kepulauan Seribu Selatan | 1 | 3,05 | 12053,41 | 0,75 | 0,75 |
| Kepulauan Seribu Utara | 4 | 5,65 | 15299,29 | 1,17 | 2,83 |

Tabel 2 di atas menampilkan *dataset* yang sudah diolah, *dataset* tersebut berisi beberapa atribut. Pada atribut *rw terdampak banjir* dan *rt terdampak banjir* merupakan hasil rata-rata dari kejadian banjir di DKI Jakarta selama tahun 2020 berdasarkan kecamatannya. *Dataset* di atas sudah melalui proses seleksi *feature*, atribut kota administrasi dihapus sehingga menyisakan 6 atribut seperti pada tabel 2.

Tabel 3. *Dataset* Hasil Normalisasi *Min-max*

| kecamatan | penderita _dbd | luas_wila yah | kepadatan_ penduduk | rw_terdampak_ banjir | rt_terdampak_ banjir |
|--------------------------|-------------------|------------------|------------------------|-------------------------|-------------------------|
| Cempaka Putih | 0,130548 | 0,058396 | 0,103953 | 0,0536 | 0,023555 |
| Gambir | 0,046997 | 0,130827 | 0,161861 | 0,0000 | 0,000000 |
| Johar Baru | 0,099217 | 0,000000 | 0,491459 | 0,0536 | 0,023555 |
| Kemayoran | 0,237598 | 0,122306 | 0,564002 | 0,1672 | 0,078518 |
| Menteng | 0,096606 | 0,104261 | 0,113030 | 0,0336 | 0,013191 |
| ... | ... | ... | ... | ... | ... |
| Pasar Rebo | 0,378590 | 0,265664 | 0,135218 | 0,1400 | 0,060302 |
| Pulo Gadung | 0,292428 | 0,331579 | 0,247252 | 0,2736 | 0,238379 |
| Kepulauan Seribu Selatan | 0,000000 | 0,017043 | 0,000000 | 0,0536 | 0,021043 |
| Kepulauan Seribu Utara | 0,007833 | 0,082206 | 0,005597 | 0,0872 | 0,086369 |

Tabel 3 di atas menampilkan *dataset* hasil dari normalisasi *min-max*. Proses normalisasi *min-max* dilakukan dengan menggunakan perintah *python*, yaitu *MinMaxScaler()* dari *sklearn library*. Normalisasi *min-max* menghasilkan nilai dengan rentang 0 sampai dengan 1, hal ini dapat dilihat pada tabel 3 di atas.

3.4. Pemodelan (*Modeling*)

Proses selanjutnya adalah dengan menerapkan algoritma *k-means* pada *dataset* yang sudah melalui hasil normalisasi. Pengolahan data ini menggunakan bahasa pemrograman *python* pada *Jupyter Notebook*. Langkah awal yang dilakukan adalah menentukan nilai *K*, nilai *K* tersebut merupakan banyaknya *cluster* yang akan dibagi. Nilai *K=2* digunakan pada penelitian ini sebagai jumlah banyaknya pembagian *cluster*. Hasil *clustering* dengan nilai *K=2* dapat dilihat pada gambar 2 berikut ini.

```
kmeans.labels_
array([0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, 1, 1, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0])
```

Gambar 2. Hasil *Clustering*

Gambar 2 menunjukkan hasil *clustering* dengan menggunakan perintah *python*, hasil *clustering* tersebut menggunakan nilai *K=2*. Pada gambar 2 di atas menunjukkan bahwa anggota *cluster* 0 lebih banyak dibandingkan anggota *cluster* 1.

3.5. Evaluasi (*Evaluation*)

Tahap selanjutnya yang dilakukan adalah evaluasi dengan menggunakan *silhouette coefficient*.

```
Pada n_clusters=2, silhouette score-nya adalah 0.41641630429260135
Pada n_clusters=3, silhouette score-nya adalah 0.37107277440791775
Pada n_clusters=4, silhouette score-nya adalah 0.25983641855061734
Pada n_clusters=5, silhouette score-nya adalah 0.2816854635044559
Pada n_clusters=6, silhouette score-nya adalah 0.28185167701920333
Pada n_clusters=7, silhouette score-nya adalah 0.27469817732635754
Pada n_clusters=8, silhouette score-nya adalah 0.2571423897636587
```

Gambar 3. *Silhouette Score*

Tahapan ini dilakukan guna melakukan uji hasil terhadap hasil *clustering* yang sudah didapatkan dengan menerapkan algoritma *k-means*. Hasil *clustering* yang baik ditunjukkan dengan *silhouette score* yang mendekati nilai 1. Adapun *silhouette score* yang dihasilkan dapat dilihat pada gambar 3 yang menunjukkan hasil *silhouette score*, Nilai *K* yang baik ditunjukkan pada nilai

K=2 sebesar 0,41642. Adapun kriteria dari *silhouette coefficient* dapat dilihat pada tabel 4 berikut.

Tabel 4. Kriteria *Silhouette Coefficient* [11]

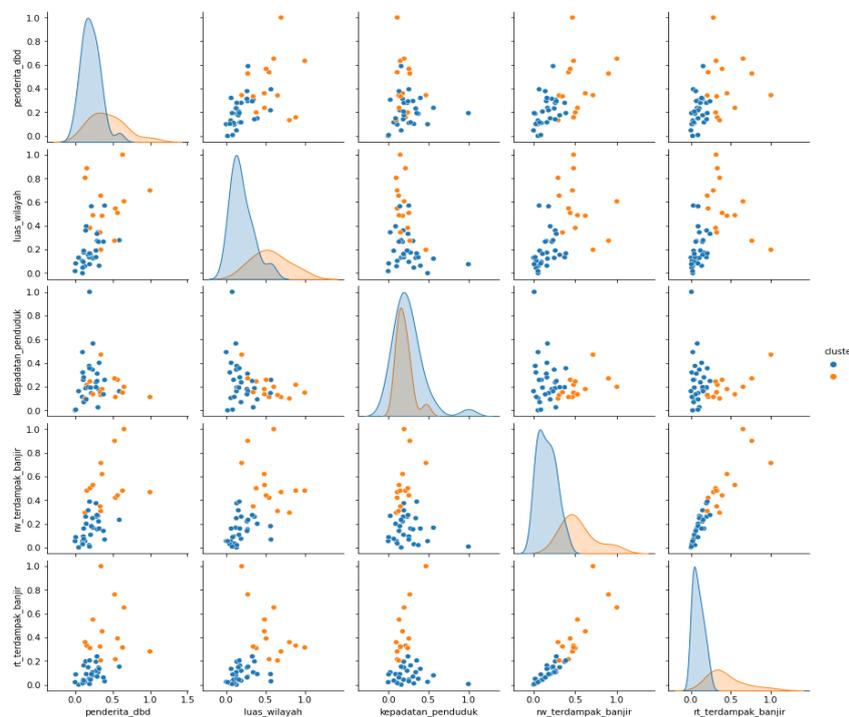
| Nilai SC | Kriteria |
|-------------|----------------|
| 0,71 - 1,00 | Struktur Kuat |
| 0,51 - 0,70 | Struktur Baik |
| 0,26 - 0,50 | Struktur Lemah |
| ≤ 0,25 | Stuktur Buruk |

Berdasarkan Tabel 4, dapat disimpulkan bahwa hasil *clustering* pada penelitian ini mempunyai nilai struktur yang lemah dengan nilai *silhouette* sebesar 0,41642.

3.6. Penyebaran (*Deployment*)

Kegiatan pada tahapan ini adalah dengan merepresentasikan mengenai hasil informasi yang didapatkan setelah melalui proses *data mining* ini. Berdasarkan proses yang sudah dilakukan dengan menerapkan algoritma *k-means*, nilai K optimal yang dapat digunakan adalah K=2. *Cluster* dibagi menjadi 2, yaitu sporadis dan endemis. C0 (sporadis) merupakan daerah dengan potensi persebaran demam berdarah tingkat sedang, sedangkan C1 (endemis) merupakan daerah dengan potensi persebaran demam berdarah tingkat tinggi.

Kegiatan selanjutnya ialah untuk melihat gambaran keterkaitan antar variabel yang digunakan. Gambaran tersebut menggunakan visualisasi *seaborn*, visualisasi tersebut mampu memvisualisasikan keterkaitan antar variabel secara bersamaan. Visualisasi tersebut dapat dilihat pada gambar 4.



Gambar 4. Visualisasi Antar Variabel

Dalam gambar 4 tersebut menampilkan kaitan antar variabel serta memperlihatkan sebaran anggota *cluster* berdasarkan variabelnya. *Cluster* 0 (sporadis) direpresentasikan dengan plot berwarna biru, sedangkan warna oranye merepresentasikan *cluster* 1 (endemis). Berikut ini merupakan tabel hasil *clustering* daerah potensi persebaran demam berdarah di DKI Jakarta yang dapat dilihat pada tabel 5.

Tabel 5. Anggota *Cluster*

| Kota Anministrasi | Kecamatan | Cluster |
|----------------------------|--------------------------|---------|
| Jakarta Pusat | Cempaka Putih | 0 |
| Jakarta Pusat | Gambir | 0 |
| Jakarta Pusat | Johar Baru | 0 |
| Jakarta Pusat | Kemayoran | 0 |
| Jakarta Pusat | Menteng | 0 |
| Jakarta Pusat | Sawah Besar | 0 |
| Jakarta Pusat | Senen | 0 |
| Jakarta Pusat | Tanah Abang | 0 |
| Jakarta Utara | Cilincing | 1 |
| Jakarta Utara | Kelapa Gading | 0 |
| Jakarta Utara | Koja | 0 |
| Jakarta Utara | Pademangan | 0 |
| Jakarta Utara | Penjaringan | 1 |
| Jakarta Utara | Tanjung Priok | 0 |
| Jakarta Barat | Cengkareng | 1 |
| Jakarta Barat | Grogol Petamburan | 0 |
| Jakarta Barat | Taman Sari | 0 |
| Jakarta Barat | Tambora | 0 |
| Jakarta Barat | Kebon Jeruk | 1 |
| Jakarta Barat | Kalideres | 1 |
| Jakarta Barat | Palmerah | 0 |
| Jakarta Barat | Kembangan | 1 |
| Jakarta Selatan | Cilandak | 0 |
| Jakarta Selatan | Jagakarsa | 0 |
| Jakarta Selatan | Kebayoran Baru | 0 |
| Jakarta Selatan | Kebayoran Lama | 0 |
| Jakarta Selatan | Mampang Prapatan | 0 |
| Jakarta Selatan | Pancoran | 0 |
| Jakarta Selatan | Pasar Minggu | 1 |
| Jakarta Selatan | Pesanggrahan | 0 |
| Jakarta Selatan | Setiabudi | 0 |
| Jakarta Selatan | Tebet | 0 |
| Jakarta Timur | Cakung | 1 |
| Jakarta Timur | Cipayung | 1 |
| Jakarta Timur | Ciracas | 1 |
| Jakarta Timur | Duren Sawit | 1 |
| Jakarta Timur | Jatinegara | 1 |
| Jakarta Timur | Kramat Jati | 1 |
| Jakarta Timur | Makasar | 0 |
| Jakarta Timur | Matraman | 0 |
| Jakarta Timur | Pasar Rebo | 0 |
| Jakarta Timur | Pulo Gadung | 0 |
| Kabupaten Kepulauan Seribu | Kepulauan Seribu Utara | 0 |
| Kabupaten Kepulauan Seribu | Kepulauan Seribu Selatan | 0 |

Kegiatan selanjutnya adalah memvisualisasikan hasil *clustering* yang sudah didapatkan dalam bentuk *maps*. *Maps* daerah potensi persebaran DBD di DKI Jakarta dapat dilihat pada gambar 5 yang menampilkan *maps cluster* daerah potensi persebaran demam berdarah di DKI Jakarta. Pada *maps* tersebut, terlihat bahwa daerah sporadis mendominasi dibandingkan dengan daerah endemis yang direpresentasikan dengan warna merah. Berdasarkan *maps* tersebut, terlihat bahwa daerah endemis atau daerah rawan DBD yang paling banyak ialah daerah Jakarta Timur, yakni tersebar di kecamatan kecamatan Cakung, Cipayung, Ciracas, Duren Sawit, Jatinegara, Kramat Jati, dan Makasar.



Gambar 5. *Maps Cluster* Daerah Potensi Persebaran DBD di DKI Jakarta

Dalam *maps* tersebut juga menunjukkan bahwa Kabupaten Kepulauan Seribu merupakan daerah sporadis, maka tingkat potensi persebaran demam berdarah di daerah tersebut sedang, hal tersebut sejalan dengan angka kasus demam berdarah yang rendah di daerah tersebut.

4. KESIMPULAN

Berdasarkan penelitian yang sudah dilakukan, dapat dikatakan bahwa algoritma *K-Means* dapat diimplementasikan guna melakukan *clustering* daerah potensi persebaran demam berdarah *dengue* di DKI Jakarta. Berdasarkan nilai *Silhouette Score* yang didapatkan, nilai *K* optimal yang dapat diimplementasikan adalah $K=2$ dengan *score* sebanyak 0,41642 serta nilai struktur lemah. Penelitian ini menghasilkan C0 (sporadis) dan C1 (endemis). Sporadis merupakan daerah dengan tingkat penyebaran demam berdarah yang sedang, sedangkan endemis merupakan daerah dengan tingkat penyebaran demam berdarah yang tinggi.

Daerah sporadis mencakup 30 kecamatan, di antaranya Gambir, Menteng, Tanah Abang, Cempaka Putih, Senen, Johar Baru, Sawah Besar, Kemayoran, Pasar Rebo, Pulo Gadung, Matraman, Palmerah, Tambora, Taman Sari, Grogol Petamburan, Koja, Kelapa Gading, Pademangan, Tanjung Priok, Pesanggrahan, Mampang Prapatan, Kebayoran Baru, Tebet, Kebayoran Lama, Cilandak, Pancoran, Jagakarsa, Setiabudi, Kepulauan Seribu Selatan, dan Kepulauan Seribu Utara. Daerah endemis di DKI Jakarta menyebar pada 14 kecamatan, di antaranya Cilincing, Penjaringan, Cengkareng, Kalideres, Kebon Jeruk, Kembangan, Pasar Minggu, Ciracas, Cakung, Makasar, Duren Sawit, Kramat Jati, Jatinegara, dan Cipayung. Berdasarkan *clustering* tersebut, diharapkan dapat membantu pemerintah dalam mengambil keputusan untuk meningkatkan upaya-upaya pencegahan dalam penyebaran demam berdarah di DKI Jakarta khususnya pada daerah endemis dengan tingkat potensi penyebarannya yang tinggi. Hal ini juga diharapkan dapat membantu masyarakat setempat guna mengetahui tingkat potensi persebaran di daerah masing-masing, sehingga dapat bekerja sama dengan pihak pemerintah untuk menekan angka penyebaran demam berdarah di DKI Jakarta.

5. SARAN

Penelitian selanjutnya perlu adanya pengembangan yang dilakukan, pengembangan tersebut dapat berupa peningkatan jumlah data yang digunakan pada penelitian, selain itu pengembangan dapat dilakukan dengan melakukan *clustering* berdasarkan objek lain selain kecamatan. Pengembangan lainnya yang dapat dilakukan peneliti selanjutnya ialah dengan menambahkan variabel lainnya yang lebih spesifik, sehingga *clustering* yang dihasilkan lebih akurat, optimal, serta dapat mencapai penelitian dengan kriteria *silhouette score* struktur yang kuat.

DAFTAR PUSTAKA

- [1] F. Nugraha, B. Haryanto, R. A. Wulandari, dan T. T. Pakasi, “Studi Ekologi Hubungan Kejadian Demam Berdarah Dengue (DBD) dengan Faktor Iklim di Kota Administrasi Jakarta Pusat, Indonesia Tahun 1999-2018,” *Jurnal Ilmu Kesehatan Masyarakat*, vol. 10, no. 03, hal. 142–148, Sep 2021, doi: 10.33221/jikm.v10i03.923.
- [2] S. Susilawati, “Dampak Perubahan Iklim Terhadap Kesehatan,” *Electronic Journal Scientific of Environmental Health And Disease*, vol. 2, no. 1, hal. 25–31, Jun 2021, doi: 10.22437/esehad.v2i1.13749.
- [3] A. I. Widyatami dan D. A. Suryawan, “Pengelompokan Daerah Rawan Demam Berdarah Dengue di Provinsi DKI Jakarta,” *Indonesian of Health Information Management Journal (INOHIM)*, vol. 9, no. 1, hal. 73–82, 2021, doi: 10.47007/inohim.v9i1.241.
- [4] D. Komaling, O. J. Sumampouw, R. C. Sondakh, F. Kesehatan, M. Universitas, dan S. Ratulangi, “Determinan kejadian demam berdarah dengue di Kabupaten Minahasa Selatan Tahun 2016-2018,” *Journal of public health and community medicine*, vol. 1, no. 1, hal. 57–64, 2020.
- [5] Kementerian Kesehatan Republik Indonesia, “Hingga Juli, Kasus DBD di Indonesia Capai 71 Ribu,” *Kementerian Kesehatan Republik Indonesia*, 2020. <https://www.kemkes.go.id/article/view/20070900004/hingga-juli-kasus-dbd-di-indonesia-capai-71-ribu.html> (diakses Nov 02, 2021).
- [6] DPRD Provinsi DKI Jakarta, “Selain Corona, Dinkes Didorong Waspada pada Sebaran DBD,” *DPRD Provinsi DKI Jakarta*, 2020. <https://dprd-dkijakartaprov.go.id/selain-corona-dinkes-didorong-waspada-pada-sebaran-dbd/> (diakses Mei 24, 2022).
- [7] C. Yuan dan H. Yang, “Research on K-Value Selection Method of K-Means Clustering Algorithm,” *J*, vol. 2, no. 2, hal. 226–235, Jun 2019, doi: 10.3390/j2020016.
- [8] N. Mirantika, A. Tsamratulain, dan F. diviana Agnia, “Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Covid-19 Di Provinsi Jawa Barat,” *Jurnal Nuansa Informatika*, vol. 15, no. 2, hal. 92–98, 2021.
- [9] B. E. Adiana, I. Soesanti, dan A. E. Permanasari, “Analisis Segmentasi Pelanggan Menggunakan Kombinasi Rfm Model dan Teknik Clustering,” *Jurnal Terapan Teknologi Informasi*, vol. 2, no. 1, hal. 23–32, Apr 2018, doi: 10.21460/jutei.2018.21.76.
- [10] Y. P. Sari, A. Primajaya, dan A. S. Y. Irawan, “Implementasi Algoritma K-Means untuk Clustering Penyebaran Tuberkulosis di Kabupaten Karawang,” *INOVTEK Polbeng - Seri Informatika*, vol. 5, no. 2, hal. 229, Nov 2020, doi: 10.35314/isi.v5i2.1457.
- [11] N. Syifa dan R. N. Fahmi, “Implementasi Metode K-Means Clustering dalam Analisis Persebaran UMKM di Jawa Barat,” *JOINS (Journal of Information System)*, vol. 6, no. 2, hal. 211–220, Des 2021, doi: 10.33633/joins.v6i2.5310.