



## Analisis Komparasi Clustering Tingkat Penyebaran Demam Berdarah Dengue Dengan Algoritma K-Means dan K-Medoids

Abdussalam Amrullah<sup>1</sup>, Teguh Muhammad Prasetyo<sup>2</sup>, Betha Nurina Sari<sup>3</sup>

<sup>1,2,3</sup>Universitas Singaperbangsa Karawang

Received: 8 Agustus 2022  
Revised: 12 Agustus 2022  
Accepted: 20 Agustus 2022

### Abstract

*Disease is a disorder that occurs in the body both in form and function so that the body cannot work properly or normally. According to WHO (World Health Organization), the number of reported cases of dengue fever increased more than 8 times. In 2000 there were 505,430 cases in the world. Then in 2010, the number of cases jumped to 2.4 million cases. In 2019, 5.2 million cases were reported. Clustering is a method in data mining that aims to group objects into clusters that have similar characteristics to one another. Cluster analysis is an important thing to do to monitor the level of spread of dengue disease so that the government can take action to deal with dengue disease in a focused manner in areas that have a high level of spread. This study utilizes data on the number of cases of dengue hemorrhagic fever and the population in Karawang. The design of the KDD method uses the K-Means and K-Medoids algorithms. Based on the evaluation results, scheme 3, which uses the original dataset or without normalization and uses the K-Means algorithm, is the best scheme with a DBI value of 0.472. Based on the evaluation results, K-Means has a better performance than K-Medoids using either the unnormalized dataset or the normalized dataset.*

**Keywords:** Data Mining, Clustering, Dengue Disease, K-Means, K-Medoids

(\* Corresponding Author: [abdussalam.amrullah18179@student.unsika.ac.id](mailto:abdussalam.amrullah18179@student.unsika.ac.id),  
[teguh.prasetyo18139@student.unsika.ac.id](mailto:teguh.prasetyo18139@student.unsika.ac.id),  
[betha.nurina@staff.unsika.ac.id](mailto:betha.nurina@staff.unsika.ac.id)

**How to Cite:** Amrullah, A., Prasetyo, T., & Sari, B. (2022). Analisis Komparasi Clustering Tingkat Penyebaran Demam Berdarah Dengue Dengan Algoritma K-Means Dan K-Medoids. *Jurnal Ilmiah Wahana Pendidikan*, 8(16), 1-10. <https://doi.org/10.5281/zenodo.7058882>.

## PENDAHULUAN

Penyakit merupakan gangguan yang terjadi pada tubuh baik berupa bentuk maupun fungsinya sehingga tubuh tidak dapat bekerja dengan semestinya atau normal. Tidak sedikit juga penyakit yang bahkan dapat menyebabkan kematian. Sehingga penyakit masih menjadi momok yang menakutkan bagi sebagian masyarakat tetapi masih ada pula yang masih menyepelekan. Salah satu penyakit yang dapat menyebabkan kematian yaitu demam berdarah dengue (DBD). Hal ini disebabkan karena gejala awal yang dimiliki oleh penderita adalah demam, nyeri pada otot dan sendi seperti demam biasa pada umumnya. Hal tersebut terjadi pada dua hari pertama, yang semakin lama akan semakin memburuk. Setelah demam tinggi selama 2-7 hari tanpa sebab yang jelas, disertai manifestasi seperti petekie, epistaxis kadang disertai muntah darah, berak darah, kesadaran menurun, bertandensi menimbulkan *shock* dan kematian. Penyakit DBD itu sendiri disebabkan oleh virus *Dengue* dan yang sering menjadi vektor penyakit tersebut nyamuk *Aedes Aegypti* (Sembiring, 2021).

Menurut WHO (*World Health Organization*), Peningkatan jumlah kasus DBD yang dilaporkan meningkat lebih dari 8 kali lipat. Pada tahun 2000 terdapat 505.430 kasus di dunia. Kemudian pada tahun 2010 angka kasus tersebut melonjak



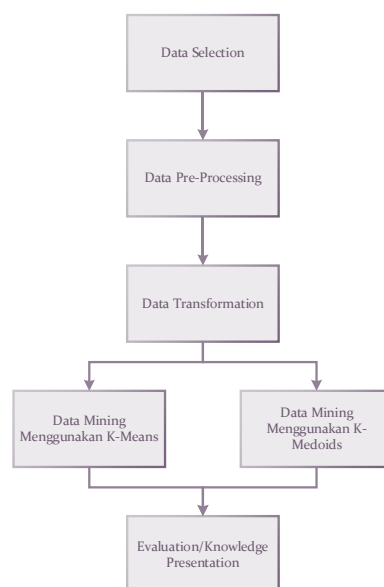
menjadi 2,4 juta kasus. Pada tahun 2019 sebanyak 5,2 juta kasus dilaporkan. Di Indonesia, kasus tertinggi terletak pada tahun 2016 yaitu sebanyak 204.171 kasus demam berdarah dengue (DBD). Angka tersebut mengalami penurunan pada tahun 2017 dan 2018 yaitu sebanyak 68.407 dan 65.602 kasus. Pelonjakan kembali terjadi pada tahun 2019 yaitu sebesar 138.127 kasus. Kota Bandung, kota Depok, dan kabupaten Bogor menjadi kota/kabupaten dengan kasus DBD tertinggi di Indonesia.

*Clustering* merupakan suatu metode pada *data mining* yang bertujuan untuk mengelompokkan objek-objek kedalam *cluster-cluster* yang memiliki kemiripan karakteristik satu dengan yang lainnya di dalam *cluster* yang sama dan memiliki perbedaan karakteristik pada objek yang berada pada *cluster* lain (Hariyanto & Shita, 2018).

Analisis *cluster* menjadi hal yang penting dilakukan untuk memantau tingkat penyebaran penyakit DBD sehingga pemerintah dapat mengambil tindakan penanganan dalam mengatasi penyakit DBD secara terfokus pada daerah yang memiliki tingkat penyebaran yang tinggi dan pencegahan pada daerah yang masih memiliki tingkat penyebaran yang rendah. Selain itu pemilihan metode dalam melakukan penelitian analisis *cluster* merupakan salah satu yang penting. Penelitian yang telah dilakukan sebelumnya menggunakan dataset yang sama yaitu pengklasteran demam berdarah dengue menggunakan algoritma K-Medoids dan menghasilkan nilai indeks *Silhouette Coefficient* sebesar 0,787 (Agustian & Dermawan, 2022) .

## METODOLOGI PENELITIAN

Penelitian ini menggunakan metode penelitian yang biasa digunakan pada proses *data mining* yaitu KDD (*Knowledge Discovery in Database*). Penelitian ini memanfaatkan data jumlah kasus demam berdarah dengue dan jumlah penduduk yang ada di Kabupaten Karawang. Adapun rancangan metode KDD menggunakan algoritma K-Means dan K-Medoids seperti pada Gambar 1.



Gambar 1. Rancangan Penelitian

### 1. Data Selection

Pada proses ini dilakukan pemilihan data dan pengecekan data pada dataset yang digunakan yaitu data kasus penyakit demam berdarah dengue dan jumlah penduduk di Kabupaten Karawang.

### 2. Data Pre-Processing

Setelah dilakukan pengecekan, kemudian dilakukan pembersihan pada data kasus penyakit DBD dan jumlah penduduk di Kabupaten Karawang dengan cara membersihkan duplikasi data, *missing value*, dan data yang tidak konsisten.

### 3. Data Transformation

Pada tahap *data transformation*, data akan dirubah menjadi bentuk yang dapat digunakan oleh yang *tools* sehingga data digunakan dalam proses *data mining*. Kedua dataset tersebut digabungkan menjadi satu sehingga dapat dilakukan proses *data mining*.

### 4. Data Mining

Pada tahapan *data mining*, pengolahan data dilakukan menggunakan algoritma *K-Means* dan *K-Medoids*. Penelitian ini melakukan 4 skema yaitu mengolah data yang sudah dinormalisasi menggunakan algoritma *K-Means* maupun *K-Medoids* dan mengolah data yang tidak dinormalisasi baik menggunakan algoritma *K-Means* maupun *K-Medoids*.

### 5. Evaluation/Knowledge Interpretation

Pada tahap ini hasil *data mining* akan dievaluasi dan diinterpretasikan sehingga dapat dijadikan sebagai informasi. Hasil *data mining* akan dievaluasi menggunakan *Davies-Bouldin Index*. Kemudian akan dilakukan penarikan kesimpulan berdasarkan hasil analisis dan evaluasi.

## HASIL DAN PEMBAHASAN

Berikut merupakan hasil dan pembahasan dari penelitian *clustering* penyakit demam berdarah dengue menggunakan algoritma *K-Means* dan *K-Medoids*.

### 1. Data Selection

Dataset yang digunakan dalam penelitian ini adalah dataset jumlah kasus penyakit demam berdarah dengue dan jumlah penduduk di Kabupaten Karawang. Atribut yang yaitu jumlah penduduk dan jumlah kasus DBD yang ada di 30 kecamatan di Kabupaten Karawang pada tahun 2020. Data tersebut diperoleh dari situs Open Data Karawang.

Tabel 1. Dataset Jumlah Kasus DBD di Kabupaten Karawang

No.	Kecamatan	Jumlah Kasus DBD
1	Banyusari	11
2	Batujaya	3
3	Ciampel	16
...	...	...
16	Lemah Abang	11
...	...	...
30	Tirtamulya	35

Tabel 2. Dataset Jumlah Penduduk di Kabupaten Karawang

No.	Kecamatan	Jumlah Penduduk
-----	-----------	-----------------

1	Banyusari	55425
2	Batujaya	77966
3	Ciampel	44104
...	...	...
16	Lemah Abang	64711
...	...	...
30	Tirtamulya	49599

## 2. Data Pre-Processing

Selanjutnya pada tahap *pre-processing* ini, data yang telah dipilih dari proses *data selection* akan dilakukan pembersihan data yang hilang (*missing value*), duplikasi data, dan data yang tidak konsisten. Pada kedua dataset yang tidak ditemukan *missing value*, duplikasi data maupun data yang tidak konsisten.

## 3. Data Transformation

Pada tahap *data transformation*, *dataset* akan digabungkan menjadi satu dataset. *Dataset* yang telah digabungkan dapat dilihat pada Tabel 3.

Tabel 3. Dataset yang Telah Digabungkan

No.	Kecamatan	Jumlah Kasus DBD	Jumlah Penduduk
1	Banyusari	11	55425
2	Batujaya	3	77966
3	Ciampel	16	44104
...	...	...	...
16	Lemah Abang	11	64711
...	...	...	...
30	Tirtamulya	35	49599

Setelah dilakukan penggabungan *dataset*, *dataset* akan dilakukan normalisasi. Normalisasi digunakan untuk membandingkan *clustering* mana memiliki hasil yang lebih baik antara *dataset* asli atau *dataset* normalisasi. Normalisasi yang dilakukan menggunakan fitur *Z-Transformation* pada Rapidminer. *Dataset* hasil normalisasi dapat dilihat pada Tabel 4.

Tabel 4. Dataset Normalisasi Jumlah Kasus di Kabupaten Karawang

No.	Kecamatan	Jumlah Kasus DBD	Jumlah Penduduk
1	Banyusari	-0.525	-0.609
2	Batujaya	-0.651	-0.019
3	Ciampel	-0.446	-0.905
...	...	...	...
16	Lemah Abang	-0.525	-0.366
...	...	...	...
30	Tirtamulya	-0.147	-0.762

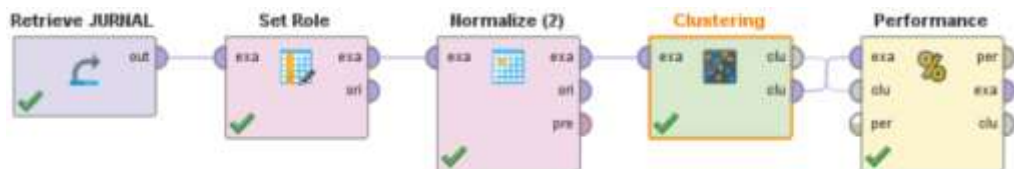
## 4. Data Mining

Proses *data mining* dilakukan menggunakan algoritma *K-Means* dan *K-Medoids* baik *dataset* asli maupun *dataset* normalisasi. Jumlah *cluster* yang

digunakan pada penelitian ini adalah 3. Proses *data mining* dilakukan dengan 4 skema yaitu :

1) Skema 1

Skema 1 menggunakan *dataset* normalisasi yang kemudian diolah menggunakan algoritma *K-Means*. Adapun operator yang digunakan untuk pengolahan data pada Rapidminer dapat dilihat Gambar 2.



Gambar 2. Operator *Data Mining* Skema 1 pada Rapidminer

Hasil *clustering* skema 1 diperoleh *Cluster 1* (C1) terdiri dari 11 kecamatan, *Cluster 2* (C2) terdiri dari 13 kecamatan dan *Cluster 3* (C3) terdiri dari 6 kecamatan. Hasil *clustering* skema 1 dapat dilihat pada Tabel 5.

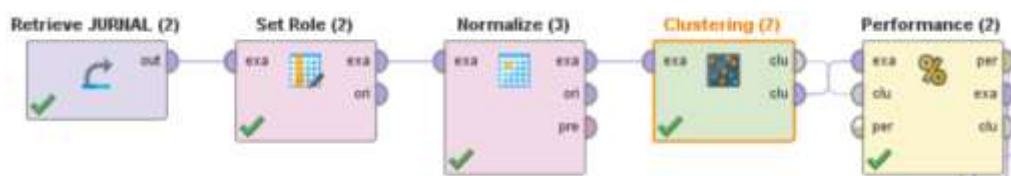
Tabel 5. Hasil *Clustering* Skema 1

Nama Kecamatan	Hasil
Banyusari	C1
Batujaya	C2
Ciampel	C1
Cibuaya	C1
Cikampek	C3
Cilamaya Kulon	C2
Cilamaya Wetan	C2
Cilebar	C1
Jatisari	C2
Jayakarta	C2
Karawang Barat	C3
Karawang Timur	C3
Klari	C3
Kotabaru	C3
Kutawaluya	C1
Lemah Abang	C2
Majalaya	C2
Pakisjaya	C1
Pangkalan	C1
Pedes	C2
Purwasari	C2
Rawamerta	C1
Rengasdengklok	C2
Tegalwaru	C1
Telagasari	C2
Telukjambe Barat	C1

<b>Telukjambe Timur</b>	C3
<b>Tempuran</b>	C2
<b>Tirtajaya</b>	C2
<b>Tirtamulya</b>	C1

## 2) Skema 2

Skema 2 menggunakan *dataset* normalisasi yang kemudian diolah menggunakan algoritma *K-Medoids*. Adapun operator yang digunakan untuk pengolahan data pada Rapidminer dapat dilihat Gambar 3.

Gambar 3. Operator *Data Mining* Skema 2 pada Rapidminer

Hasil *clustering* skema 2 diperoleh *Cluster 1* (C1) terdiri dari 12 kecamatan, *Cluster 2* (C2) terdiri dari 12 kecamatan dan *Cluster 3* (C3) terdiri dari 6 kecamatan. Hasil *clustering* skema 2 dapat dilihat pada Tabel 6.

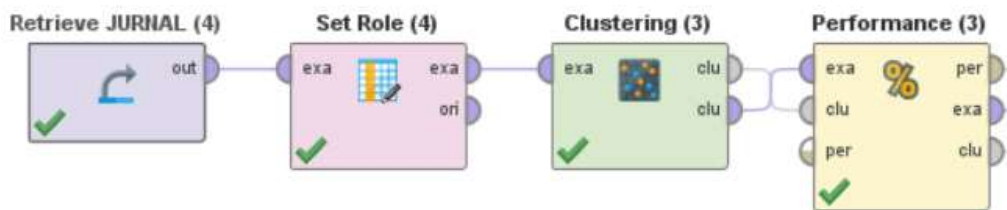
Tabel 6. Hasil *Clustering* Skema 2

<b>Nama Kecamatan</b>	<b>Hasil</b>
<b>Banyusari</b>	C2
<b>Batujaya</b>	C1
<b>Ciampel</b>	C2
<b>Cibuaya</b>	C1
<b>Cikampek</b>	C3
<b>Cilamaya Kulon</b>	C1
<b>Cilamaya Wetan</b>	C1
<b>Cilebar</b>	C2
<b>Jatisari</b>	C1
<b>Jayakarta</b>	C1
<b>Karawang Barat</b>	C3
<b>Karawang Timur</b>	C3
<b>Klari</b>	C3
<b>Kotabaru</b>	C3
<b>Kutawaluya</b>	C1
<b>Lemah Abang</b>	C1
<b>Majalaya</b>	C2
<b>Pakisjaya</b>	C2
<b>Pangkalan</b>	C2
<b>Pedes</b>	C1
<b>Purwasari</b>	C2
<b>Rawamerta</b>	C2
<b>Rengasdengklok</b>	C1

<b>Tegalwaru</b>	C2
<b>Telagasari</b>	C2
<b>Telukjambe Barat</b>	C2
<b>Telukjambe Timur</b>	C3
<b>Tempuran</b>	C1
<b>Tirtajaya</b>	C1
<b>Tirtamulya</b>	C2

## 3) Skema 3

Skema 3 menggunakan *dataset* asli yang kemudian diolah menggunakan algoritma *K-Means*. Adapun operator yang digunakan untuk pengolahan data pada Rapidminer dapat dilihat Gambar 4.

Gambar 4. Operator *Data Mining* Skema 3 pada Rapidminer

Hasil *clustering* skema 3 diperoleh *Cluster 1* (C1) terdiri dari 13 kecamatan, *Cluster 2* (C2) terdiri dari 10 kecamatan dan *Cluster 3* (C3) terdiri dari 7 kecamatan. Hasil *clustering* skema 2 dapat dilihat pada Tabel 7.

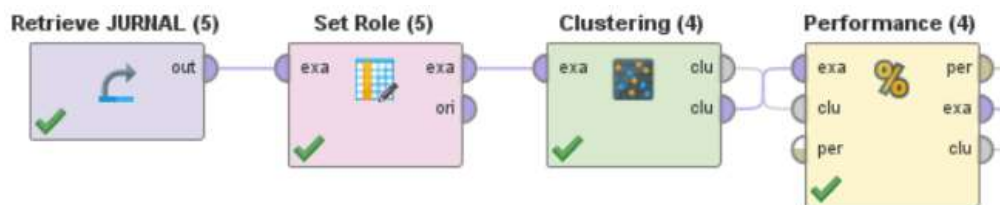
Tabel 7. Hasil *Clustering* Skema 3

<b>Nama Kecamatan</b>	<b>Hasil</b>
<b>Banyusari</b>	C2
<b>Batujaya</b>	C1
<b>Ciampel</b>	C2
<b>Cibuaya</b>	C2
<b>Cikampek</b>	C3
<b>Cilamaya Kulon</b>	C1
<b>Cilamaya Wetan</b>	C1
<b>Cilebar</b>	C2
<b>Jatisari</b>	C1
<b>Jayakarta</b>	C1
<b>Karawang Barat</b>	C3
<b>Karawang Timur</b>	C3
<b>Klari</b>	C3
<b>Kotabaru</b>	C3
<b>Kutawaluya</b>	C1
<b>Lemah Abang</b>	C1
<b>Majalaya</b>	C1
<b>Pakisjaya</b>	C2
<b>Pangkalan</b>	C2

<b>Pedes</b>	C1
<b>Purwasari</b>	C1
<b>Rawamerta</b>	C2
<b>Rengasdengklok</b>	C3
<b>Tegalwaru</b>	C2
<b>Telagasari</b>	C1
<b>Telukjambe Barat</b>	C2
<b>Telukjambe Timur</b>	C3
<b>Tempuran</b>	C1
<b>Tirtajaya</b>	C1
<b>Tirtamulya</b>	C2

## 4) Skema 4

Skema 4 menggunakan *dataset* asli yang kemudian diolah menggunakan algoritma *K-Medoids*. Adapun operator yang digunakan untuk pengolahan data pada Rapidminer dapat dilihat Gambar 5.



Gambar 5. Operator *Data Mining* Skema 4 pada Rapidminer

Hasil *clustering* skema 4 diperoleh *Cluster 1* (C1) terdiri dari 12 kecamatan, *Cluster 2* (C2) terdiri dari 11 kecamatan dan *Cluster 3* (C3) terdiri dari 7 kecamatan. Hasil *clustering* skema 2 dapat dilihat pada Tabel 8.

Tabel 8. Hasil *Clustering* Skema 4

<b>Nama Kecamatan</b>	<b>Hasil</b>
<b>Banyusari</b>	C2
<b>Batujaya</b>	C1
<b>Ciampel</b>	C2
<b>Cibuaya</b>	C2
<b>Cikampek</b>	C3
<b>Cilamaya Kulon</b>	C1
<b>Cilamaya Wetan</b>	C1
<b>Cilebar</b>	C2
<b>Jatisari</b>	C1
<b>Jayakarta</b>	C1
<b>Karawang Barat</b>	C3
<b>Karawang Timur</b>	C3
<b>Klari</b>	C3
<b>Kotabaru</b>	C3
<b>Kutawaluya</b>	C2

<b>Lemah Abang</b>	C1
<b>Majalaya</b>	C1
<b>Pakisjaya</b>	C2
<b>Pangkalan</b>	C2
<b>Pedes</b>	C1
<b>Purwasari</b>	C1
<b>Rawamerta</b>	C2
<b>Rengasdengklok</b>	C3
<b>Tegalwaru</b>	C2
<b>Telagasari</b>	C1
<b>Telukjambe Barat</b>	C2
<b>Telukjambe Timur</b>	C3
<b>Tempuran</b>	C1
<b>Tirtajaya</b>	C1
<b>Tirtamulya</b>	C2

### 5. Evaluation/Knowledge Presentation

Tahap ini merupakan tahap evaluasi hasil *clustering* sehingga dapat membandingkan performa mana yang lebih baik. Penelitian menggunakan *Davies-Bouldin Index* dalam mengevaluasi hasil *clustering*. Keempat skema dibandingkan nilai dibandingkan nilai DBInya kemudian menyimpulkan skema mana yang memiliki performa yang lebih baik. Hasil evaluasi dari keempat skema dapat dilihat pada Tabel 9.

Tabel 9. Hasil Evaluasi

<b>Algoritma</b>	<b>Dataset Asli</b>	<b>Dataset Normalisasi</b>
<b><i>K-Means</i></b>	0.472	0.794
<b><i>K-Medoids</i></b>	0.515	0.795

Berdasarkan Tabel 9 dapat disimpulkan bahwa skema 3 yang menggunakan *dataset* tanpa normalisasi dan algoritma *K-Means* lebih baik dari skema lainnya. Skema 3 memiliki nilai DBI sebesar 0.472 sedangkan skema yang terendah lainnya yaitu skema 4 memiliki nilai DBI sebesar 0.515. Rincian kriteria hasil *clustering* skema 4 dapat dilihat pada Tabel 10.

Tabel 10. Rincian Kriteria Skema 4

<b>Cluster</b>	<b>Rata- Rata Jumlah Kasus DBD</b>	<b>Rata-Rata Jumlah Penduduk</b>
<b><i>Cluster 1</i></b>	19	70056.356
<b><i>Cluster 2</i></b>	13	46915.800
<b><i>Cluster 3</i></b>	136	140161.143

Berdasarkan Tabel 10 *Cluster 1* memiliki karakteristik dengan rata-rata jumlah kasus DBD yang cukup tinggi tetapi memiliki rata-rata jumlah penduduk yang tinggi pula. Sedangkan pada *Cluster 2* memiliki karakteristik dengan rata-rata

jumlah kasus DBD tinggi tetapi memiliki rata-rata jumlah yang rendah. Dan *Cluster* 3 memiliki karakteristik yang kedua atribut sangat tinggi.

### **KESIMPULAN**

Berdasarkan penelitian yang telah dilakukan dalam *clustering* tingkat penyebaran demam berdarah dengue di Kabupaten Karawang dapat disimpulkan bahwa :

1. Berdasarkan hasil evaluasi, skema 3 yaitu menggunakan *dataset* asli atau tanpa normalisasi dan menggunakan algoritma *K-Means* merupakan skema terbaik dengan nilai DBI sebesar 0.472.
2. Hasil *clustering* terbaik yaitu *Cluster 1* memiliki karakteristik rata-rata jumlah kasus DBD yang rendah dengan rata-rata jumlah penduduk sedang. Sedangkan *Cluster 2* memiliki karakteristik rata-rata jumlah kasus DBD yang rendah dengan rata-rata jumlah penduduk yang rendah pula. *Cluster 3* memiliki karakteristik jumlah penduduk dan jumlah kasus DBD yang tinggi.
3. Berdasarkan hasil evaluasi, *K-Means* memiliki performa yang lebih baik dibandingkan *K-Medoids* baik menggunakan *dataset* tanpa normalisasi maupun *dataset* normalisasi.

### **DAFTAR PUSTAKA**

- Agustian, D. R., & Dermawan, B. A. (2022). Analisis Clustering Demam Berdarah Dengue dengan Algoritma K-Medoids (Studi Kasus Kabupaten Karawang). *Jurnal Informatika Dan Komputer (JIKO)*, 6(1), 18–26.
- 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.
- Sembiring, M. A. (2021). Penerapan Metode Algoritma K-Means Clustering Untuk Pemetaan Penyebaran Penyakit Demam Berdarah Dengue (Dbd). *Journal of Science and Social Research*, 4(3), 336. <https://doi.org/10.54314/jssr.v4i3.712>