

## PENGELOMPOKAN KASUS PENYAKIT MALARIA BERDASARKAN KABUPATEN DI PROVINSI JAWA BARAT DENGAN ALGORITMA K-MEDOIDS

Deri Nur Alamsyah, Riri Narasati, Dodi Solihudin, Ahmad Rifa'i

Teknik Informatika, STMIK IKMI Cirebon

Jl. Perjuangan No.10B, Karyamulya, Kec. Kesambi Kota Cirebon, Jawa Barat

*derinuralamsyah2001@gmail.com*

### ABSTRAK

Penyakit malaria, disebabkan oleh parasit dari genus Plasmodium melalui gigitan nyamuk Anopheles. Dampaknya melibatkan peningkatan kematian pada kelompok rentan seperti bayi, anak di bawah lima tahun, dan ibu melahirkan, sambil merugikan produktivitas kerja dan mengancam citra politik negara. Penderita utamanya berasal dari daerah pedesaan dan ekonomi lemah. Tujuan penelitian ini adalah mengidentifikasi tingkat keparahan kasus malaria berdasarkan provinsi menggunakan teknik data mining K-medoids dengan aplikasi RapidMiner. Dari hasil analisis, 24 provinsi teridentifikasi dengan tingkat penyakit malaria tertinggi, termasuk di Jawa Barat. Rekomendasi diberikan agar pemerintah memberikan perhatian khusus pada provinsi-provinsi ini untuk mendukung upaya penanggulangan penyakit malaria. Tujuan dan manfaat dari penelitian ini adalah untuk memberikan pengetahuan mengenai hasil dari analisis menggunakan metode *k-medoids* selain itu untuk memberikan kontribusi dalam mendalami pengelompokan pada kasus penyakit malaria di kabupaten berdasarkan provinsi.

**Kata kunci :** Data Mining, K – Medoids, RapidMiner, Penyakit Malaria, Anopheles, Plasmodium

### 1. PENDAHULUAN

Penyakit malaria merupakan salah satu tantangan utama dalam bidang kesehatan global. dengan dampak khususnya dirasakan di negara-negara tropis. Provinsi Jawa Barat, sebagai bagian dari wilayah yang rentan terhadap penyakit ini, tidak terkecuali dari masalah kesehatan serius yang diakibatkan oleh malaria. Seiring dengan kemajuan teknologi informasi, penerapan data mining muncul sebagai alat yang potensial untuk mengatasi kompleksitas dalam analisis penyakit menular, termasuk malaria. Penelitian ini menjawab panggilan untuk memanfaatkan kekuatan data mining dalam pemahaman dan pengelolaan kasus penyakit malaria di tingkat kabupaten di Provinsi Jawa Barat.

Provinsi Jawa Barat memiliki karakteristik geografis dan demografis yang bervariasi, menciptakan tantangan unik dalam penanganan penyakit ini. Dengan memfokuskan penelitian pada tingkat kabupaten, kami berusaha mendapatkan wawasan yang lebih mendalam tentang sebaran geografis penyakit malaria dan faktor-faktor yang mungkin memengaruhi penyebarannya. Hal ini penting untuk memberikan dasar yang kuat bagi kebijakan kesehatan yang dapat diadaptasi dan respons yang cepat terhadap perubahan epidemiologi di tingkat lokal.

Algoritma K-medoids dipilih sebagai metode analisis utama dalam penelitian ini karena kemampuannya dalam mengelompokkan data menjadi kelompok-kelompok yang signifikan berdasarkan pola yang ada [1]. Berdasarkan penelitian terdahulu dari Paper 1 yang di lakukan [2] "Pemanfaatan algoritma k-medoids untuk klustering kecambah kelapa sawit" membahas tentang Tujuan untuk menganalisis faktor internal dan eksternal yang memengaruhi pemasaran

kecambah kelapa sawit serta menentukan strategi prioritas untuk pemasaran. Mereka menggunakan algoritma pengelompokan K-medoids untuk menganalisis kecambah kelapa sawit dan menemukan bahwa algoritma tersebut berhasil mengelompokkan kecambah ke dalam tiga kategori: sangat baik, baik, dan kurang baik.

Dengan menerapkan algoritma ini pada data kasus penyakit malaria di berbagai kabupaten, diharapkan kita dapat mengidentifikasi pola-pola tertentu yang dapat menjadi dasar bagi tindakan pencegahan yang lebih terarah. Penggunaan data mining dalam konteks ini juga memungkinkan kita untuk memanfaatkan informasi yang mungkin tidak terlihat dengan jelas melalui metode analisis konvensional.

Selain itu, penelitian ini diarahkan untuk membuka pintu menuju pengembangan model prediktif yang lebih canggih, memanfaatkan data historis untuk meramalkan potensi penyebaran penyakit malaria di masa depan. Dengan begitu, penelitian ini tidak hanya mengeksplorasi kondisi saat ini tetapi juga memberikan kontribusi pada pengembangan strategi jangka panjang dalam menghadapi tantangan kesehatan masyarakat. Penelitian ini bertujuan untuk menerapkan algoritma K-Medoids dalam klusterisasi. Yang terdiri dari penerapan analisis data dan penemuan algoritma yang menghasilkan enumerasi tertentu terhadap pola pada data.

### 2. TINJAUAN PUSTAKA

#### 2.1. Data Mining

Data mining adalah istilah yang digunakan untuk menemukan informasi yang tersembunyi di database. Penambangan data adalah proses semi-otomatis yang

menggunakan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk mengekstraksi dan mengidentifikasi informasi yang berpotensi berguna dan berguna yang disimpan dalam database besar [3].

**2.2. Clustering**

Clustering adalah proses pengelompokan data ke dalam sebuah cluster, sehingga berisidata yang semirip mungkin dan berbeda dengan objek cluster lainnya [4].

**2.3. Algoritma K-medoids**

Algoritma k-medoids merupakan metode berbasis partisi yang menggunakan objek representatif yang disebut medoids sebagai titik pusat atau centroid[5].

**2.4. Davies Bouldin Index**

DBI yang diperoleh adalah 0,626 seperti terlihat pada Gambar 4 tampilan uji performance di atas. Dari nilai ini, evaluasi Cluster adalah baik, karena mendekati nol. Semakin kecil DBI, maka evaluasi terhadap Cluster dikatakan baik [6].

**2.5. Knowledge Discovery in Database**

KDD merupakan proses dari menggunakan data mining untuk mencari informasi-informasi yang berharga, pola yang ada di dalam data, yang melibatkan algoritme untuk mengidentifikasi pola pada data [7]

**2.6. Penyakit Malaria**

Malaria adalah penyakit mematikan yang menjadi masalah di berbagai negara. Metode yang paling umum untuk mendeteksi malaria adalah dengan memeriksanya secara manual, yang memakan waktu [8].

**3. METODE PENELITIAN**

Metode yang digunakan untuk memecahkan masalah yang ada bertujuan untuk menunjukkan bagaimana sebuah model data mining dapat digunakan untuk mengetahui hasil pengelompokan kabupaten dan kota melalui penerapan algoritma k-medoids. Adapun untuk tahap metode penelitian ini dimulai dengan identifikasi masalah, pengumpulan data, analisis data dengan data mining, dan melakukan evaluasi.

Data pada penelitian ini di ambil dari Open Data Jabar Badan Provinsi Jawa Barat pada tanggal 29 November 2023 yang mencakup informasi tentang Kasus Penyakit Malaria di Provinsi Jawa Barat pada tahun 2016-2022 dengan total 6 variabel. Analisis data menggunakan teknik KDD (Knowledge Discovery in Databases) untuk mengelompokan kasus penyakit malaria berdasarkan kabupaten di provinsi jawa barat dengan algoritma k-medoids dapat melibatkan beberapa langkah penting. Berikut adalah panduan langkah demi langkah untuk melakukan analisis tersebut:

**3.1. Data Selection**

Data Selection adalah proses pemilihan (sampling) data dari suatu dataset yang harus dilakukan sebelum data mining di KDD. Data pemrosesan pesan disimpan dalam file terpisah dari database aplikasi.

**3.2. Data Preprocessing**

Data preprocessing adalah teknik penambahan data. Sebelum tahap konstruksi. pertama, data mentah diproses. Preprocessing data atau data preprocessing dilakukan dengan menghilangkan pesan error.

**3.3. Transformation**

Transformasi data digunakan dalam pemrosesan data untuk mengubah data menjadi format yang sesuai [9]. Banyak teknik transformasi data termasuk normalisasi, pemilihan atribut, dan diskritisasi.

**3.4. Data Mining**

Sebuah proses di mana berbagai teknik digunakan untuk mengekstraksi pola potensial dari data yang berguna.

**3.5. Interpretation/Evaluation**

Langkah ini melibatkan pemeriksaan apakah model atau data yang dihasilkan konsisten dengan informasi atau asumsi sebelumnya.

**4. HASIL DAN PEMBAHASAN**

**4.1. Data Selection**

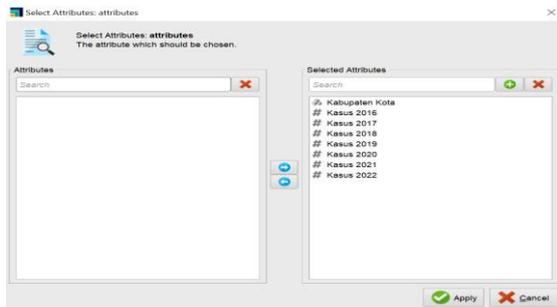
Penelitian ini menggunakan dataset yang di peroleh dari open data jabar dengan dari tahun 2016-2022 seperti pada tabel di bawah ini.

Tabel 1. Data Jawa Barat tTahun 2016-2022

no	kode_p rovinsi	nama_provinsi	kode_kabup aten_kota	nama_kabupaten_kota	jumlah_kas us	satuan	tahun
1	32	JAWA BARAT	3201	KABUPATEN BOGOR	0	ORANG	2016
2	32	JAWA BARAT	3202	KABUPATEN SUKABUMI	66	ORANG	2016
3	32	JAWA BARAT	3203	KABUPATEN CIANJUR	0	ORANG	2016
4	32	JAWA BARAT	3204	KABUPATEN BANDUNG	1	ORANG	2016
5	32	JAWA BARAT	3205	KABUPATEN GARUT	5	ORANG	2016
6	32	JAWA BARAT	3206	KABUPATEN TASIKMALAYA	59	ORANG	2016
7	32	JAWA BARAT	3207	KABUPATEN CIAMIS	0	ORANG	2016
8	32	JAWA BARAT	3208	KABUPATEN KUNINGAN	0	ORANG	2016
9	32	JAWA BARAT	3209	KABUPATEN CIREBON	11	ORANG	2016

no	kode_p rovinsi	nama_provinsi	kode_kabup aten_kota	nama_kabupaten_kota	jumlah_kas us	satuan	tahun
10	32	JAWA BARAT	3210	KABUPATEN MAJALENGKA	0	ORANG	2016
...	...	...	...	...	...	...	...
189	32	JAWA BARAT	3279	KOTA BANJAR	0	ORANG	2022

Penulis memilih data yang diperlukan, dan kemudian data tersebut disesuaikan sesuai dengan kebutuhan. Data penelitian terdiri dari 8 atribut, dengan total 27 daftar. Data yang akan digunakan dalam penelitian ini sebagai berikut.



#### 4.2. Preprocessing

Preprocessing data dilakukan, yang berarti membersihkan data yang hilang atau kosong sebelum melanjutkan proses. Ini karena data yang digunakan peneliti tidak memiliki data yang hilang atau kosong yang dapat dilihat pada gambar di bawah ini.

Gambar 1. Missing Value dataset Penelitian

#### 4.3. Transformation

Pada tahap ini, proses transformasi data dilakukan untuk menyediakan format yang tepat untuk digunakan dalam proses data mining. Tujuannya adalah untuk memudahkan koordinasi data yang diproses oleh algoritma dan alat yang digunakan dalam penelitian, yaitu rapidminer.

Tabel 2. Tahap Transformasi

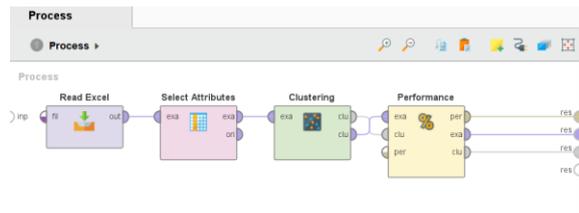
Kabupaten Kota	Kasus 2016	Kasus 2017	Kasus 2018	Kasus 2019	Kasus 2020	Kasus 2021	Kasus 2022
KABUPATEN BOGOR	0	0	7	9	6	5	11
KABUPATEN SUKABUMI	66	4	36	43	11	7	28
KABUPATEN CIANJUR	0	2	0	0	3	3	2
KABUPATEN BANDUNG	1	1	9	0	2	12	7
KABUPATEN GARUT	5	5	0	5	3	3	9
KABUPATEN TASIKMALAYA	59	33	27	20	7	15	29
KABUPATEN CIAMIS	0	3	8	3	1	2	1
KABUPATEN KUNINGAN	0	0	3	1	0	0	1
KABUPATEN CIREBON	11	9	18	5	3	3	3
...		0	0	0	0	0	2
KOTA BANJAR	0	4	0	0	0	0	0

Pada tabel diatas menunjukkan tahap transformasi, atribut periode digantikan oleh atribut nomer. Namun, karena data yang akan diproses untuk data mining sudah dalam format yang tepat di bawah ini merupakan tabel hasil transformasi.

#### 4.4. Data Mining

Langkah yang dilakukan adalah mengkonfigurasi algoritma dengan mengatur parameter seperti jumlah cluster (K). Clustering menghasilkan pengelompokan data produksi yang dapat digunakan untuk mengidentifikasi wilayah dengan tingkat produksi rendah, sedang, dan tinggi. Operator yang digunakan untuk menentukan cluster penyakit malaria di Jawa Barat antara lain adalah atribut “Read Excel” yang

digunakan untuk membaca data utama, atribut “Select Attribute” yang digunakan untuk memilih subset dari kumpulan sampel, dan lain-lain. Atribut Set Role digunakan untuk menghapus atribut dan menentukan peran atau tipe data suatu atribut dalam suatu kumpulan data. Atribut operator pengelompokan yang mengklasifikasikan data ke dalam nilai yang telah ditentukan dan fungsi kinerja yang digunakan untuk mengevaluasi kinerja model memberikan daftar tujuan kinerja tergantung pada tugas yang ditentukan. Contoh model cluster K-medoids ditunjukkan pada gambar di bawah.



Gambar 2. model cluster K-medoids

Hasil proses gambar di atas dapat di lihat pada gambar berikut.

**Cluster Model**

Description

Folder View

Cluster 0: 1 items  
Cluster 1: 26 items  
Total number of items: 27

Gambar di atas menunjukkan *cluster* yang digunakan pada dari tahun 2016 - 2022 terdiri dari tiga item. *Cluster* pertama berisi 1 elemen, *cluster* kedua berisi 26 elemen. Hasil proses cluster menghasilkan centroid seperti pada gambar di bawah.

Attribute	cluster_0	cluster_1
Kasus 2016	14	0
Kasus 2017	8	4
Kasus 2018	0	0
Kasus 2019	3	0
Kasus 2020	187	0
Kasus 2021	19	0
Kasus 2022	209	0

Gambar 3. Proses Cluster

Pada tahun 2022, terdapat 1 kota pada cluster 0, pada cluster 1 terdapat 26 kota, sehingga totalnya ada 27.

Tabel 3. Hasil Analisis Cluster Penyakit Malaria Di Provinsi Jawa Barat

Cluster 0	01.Kota Depok
Cluster 1	01.Kabupaten Bogor
	02.Kabupaten Sukabumi
	03.Kabupaten Cianjur
	04.Kabupaten Bandung
	05.Kabupaten Garut
	06.Kabupaten Tasikmalaya
	07.Kabupaten Ciamis
	08.Kabupaten Kuningan
	09.Kabupaten Cirebon
	10.Kabupaten Majalengka
	11.Kabupaten Sumedang
	12.Kabupaten Indramayu
	13.Kabupaten Subang
	14.Kabupaten Purwakarta
	15.Kabupaten Karawang
	16.Kabupaten Bekasi
	17.Kabupaten Bandung Barat

18.Kabupaten Pangandaran
19. Kota Bogor
20.Kota Sukabumi
21.Kota Bandung
22.Kota Cirebon
23.Kota Bekasi
24.Kota Cimahi
25.Kota Tasikmalaya
26.Kota Banjar

Tabel di atas menunjukkan hasil analisis cluster penyakit malaria di provinsi jawa barat.

Tabel 4. hasil Davis-Bouldin Index (DBI)

Percobaan	Jumlah Cluster	Nilai DBI
1	2	0.090
2	3	0.651
3	4	1.313
4	5	1.190

Tabel di atas menunjukkan hasil Davis-Bouldin Index (DBI) dimana tabel tersebut menunjukkan hasil untuk nilai k = 2 yaitu 0.090 maka nilai k dipilih pada percobaan kedua dan nilai k = 2 dipilih pada percobaan.

Criterion

Avg. within centroid dis...

Avg. within centroid dis...

Avg. within centroid dis...

**Davies Bouldin**

Davies Bouldin: 0.090

Davies Bouldin

Gambar 3. hasil dbi dari cluster kedua

Gambar di atas menunjukkan hasil dbi dari cluster kedua, dua cluster digunakan dalam penelitian ini karena semakin kecil nilai DBI maka cluster yang dihasilkan akan semakin optimal[10].

Setelah menerapkan algoritma k-medoids pada pengelompokan kabupaten dan kota berdasarkan penyakit malaria, diperoleh hasil cluster dengan tiga cluster pada tahun 2022. Artinya, klaster 0 memiliki 1 kota, dan cluster 1 memiliki 26 kabupaten.

Pada Cluster 0 Dicitrakan oleh rata rata pada 14 kasus di tahun 2016, 8 kasus di tahun 2017, 0 kasus di tahun 2018, 3 kasus di tahun 2019, 187 kasus di tahun 2020, 19 kasus di tahun 2021, dan 209 kasus di tahun 2022.

Dan Pada Cluster 1 Dicitrakan oleh rata rata pada 4 kasus di tahun 2017 sedangkan pada tahun 2016, 2018 dan seterusnya hanya ada pada rata-rata 0 kasus

### 5. KESIMPULAN DAN SARAN

Setelah menerapkan algoritma k-medoids pada pengelompokan kabupaten dan kota berdasarkan penyakit malaria, diperoleh hasil cluster dengan dua cluster. masing-masing cluster memiliki 0 memiliki 1 anggota, cluster 1 memiliki 26 anggota. Hasil pengelompokan menunjukkan adanya 2 cluster, yaitu

cluster 0 terdiri dari 1 kabupaten yaitu kabupaten depok, cluster 1 terdiri dari 26 kabupaten diantaranya Bogor, Sukabumi, Cianjur, Bandung, Garut, Tasikmalaya, Ciamis, Kuningan, Cirebon, Majalengka, Sumedang, Indramayu, Subang, Purwakarta, Karawang, Bekasi, Bandung Barat dan Pangandaran. 3. Cluster 0 masuk dalam Tingkat kasus malaria tertinggi dengan jumlah rata-rata 209 kasus oleh karena itu pemerintah harus memberi perhatian lebih, cluster 1 masuk dalam Tingkat malaria sedang dengan jumlah rata-rata 4 kasus malaria oleh karena itu pemerintah tetap harus memberikan perhatian untuk mengurangi kasus malaria.

#### DAFTAR PUSTAKA

- [1] N. Arminarahmah, A. G. Daengs, J. Tata Hardinata, and I. Kalimantan Muhammad Arsyad Al Banjari, "Klusterisasi Impor Beras Di Indonesia Menurut Negara Asal Utama Menggunakan Algoritma K-Medoids," *J. Ris. Sist. Inf. Dan Tek. Inform.*, vol. 8, no. 2, pp. 793–801, 2023, [Online]. Available: <https://tunasbangsa.ac.id/ejurnal/index.php/jurasik>
- [2] S. Nuraini, I. Gunawan, and W. Saputra, "Utilization of K-Medoids Algorithm for Klustering of Oil Palm Sprouts," *JOMLAI J. Mach. Learn. Artif. Intell.*, vol. 1, no. 1, pp. 11–22, 2022, doi: 10.55123/jomlai.v1i1.160.
- [3] Agung Nugraha, Odi Nurdiawan, and Gifthera Dwilestari, "Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Yana Sport," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 6, no. 2, pp. 1–7, 2022.
- [4] D. Nasution, S. Solikhun, and D. Nasution, "Penerapan K-Medoids Dalam Mengelompokkan Produksi Padi Di Indonesia Pada Masa Pandemi Covid-19," *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 4, no. 2, pp. 26–35, 2022, [Online]. Available: <https://tunasbangsa.ac.id/seminar/index.php/senaris/article/view/205%0Ahttps://tunasbangsa.ac.id/seminar/index.php/senaris/article/download/205/206>
- [5] R. N. Ibrahim and M. N. Hayati, "Penerapan Algoritma K-Medoids pada Pengelompokan Wilayah Desa atau Kelurahan di Kabupaten Kutai Kartanegara," *J. EKSPONENSIAL*, vol. 11, no. 2, pp. 153–158, 2020, [Online]. Available: <http://jurnal.fmipa.unmul.ac.id/index.php/exponensial/article/view/658>
- [6] Luth Fimawahib, Imam Rangga Bakti, and Asep Supriyanto, "Algoritma K-Medoids untuk Pengelompokan Produksi Padi dan Beras sebagai Upaya Optimalisasi Ketahanan Pangan di Provinsi Riau," *SATIN - Sains dan Teknol. Inf.*, vol. 8, no. 2, pp. 13–24, 2022, doi: 10.33372/stn.v8i2.877.
- [7] F. Amelia, I. Iskandar, S. Kurnia Gusti, and E. Haerani, "Krea-tif: jurnal teknik informatika clustering keluarga miskin desa bina baru dengan metode k-medoids," *Krea-Tif J. Tek. Inform.*, vol. 11, no. 1, pp. 1–13, 2023, doi: 10.32832/krea-tif.v11i1.14104.
- [8] Y. Yohannes, S. Devella, and K. Arianto, "Deteksi Penyakit Malaria Menggunakan Convolutional Neural Network Berbasis Saliency," *JUITA J. Inform.*, vol. 8, no. 1, p. 37, 2020, doi: 10.30595/juita.v8i1.6671.
- [9] M. Akbar and Y. Rahmanto, "Desain Data Warehouse Penjualan Menggunakan Nine Step Methodology Untuk Business Intelegency Pada Pt Bangun Mitra Makmur," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 1, no. 2, pp. 137–146, 2020, doi: 10.33365/jatika.v1i2.331.
- [10] B. E. Adiana, I. Soesanti, and A. E. Permanasari, "Analisis Segmentasi Pelanggan Menggunakan Kombinasi Rfm Model Dan Teknik Clustering," *J. Terap. Teknol. Inf.*, vol. 2, no. 1, pp. 23–32, 2018, doi: 10.21460/jutei.2018.21.76.