

Implementasi Partitioning Around Medoids Pada Visualisasi Penyebaran Penyakit DBD di Sumatera Utara

Wahyu Nur Fadillah^{1, a)}, Yulita Molliq Rangkuti^{1, b)}, dan Ichwanul Muslim Karo Karo^{1, c)}

¹Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Medan

^{a)}wahyunurfadillah313@gmail.com

^{b)}yulitamolliq@unimed.ac.id

^{c)}imkarokaro@gmail.com

Abstrak. Penelitian ini bertujuan untuk mengembangkan Sistem Informasi Geografis (SIG) menggunakan metode Partitioning Around Medoids (PAM) dalam pemetaan kasus demam berdarah dengue (DBD) di Provinsi Sumatera Utara. DBD memiliki gejala demam tinggi, pendarahan, dan tingkat kematian yang tinggi jika tidak ditangani dengan cepat. Oleh karena itu, pemetaan kasus DBD sangat penting dalam upaya penanggulangan dan pencegahan penyakit ini. Metode PAM digunakan dalam penelitian ini untuk mengelompokkan data kasus DBD berdasarkan karakteristik yang serupa. Hasil analisis menunjukkan bahwa terdapat tiga klaster yang terbentuk di Provinsi Sumatera Utara. cluster pertama terdiri dari 18 kabupaten/kota adalah cluster rendah, cluster kedua terdiri dari 3 kabupaten/kota adalah cluster tinggi, dan cluster ketiga terdiri dari 3 kabupaten/kota cluster sedang. Implementasi PAM dilakukan dengan menggunakan metode perhitungan jarak Minkowski yang mana penerapan jarak tersebut akan diuji dengan Silhouette Index pada beberapa jumlah cluster. Adapun jumlah cluster terbaik untuk implementasi PAM yang diperoleh adalah 3 cluster dengan nilai Silhouette Index 0.5275.

Kata Kunci: DBD, Partitioning Around Medoids, Minkowski

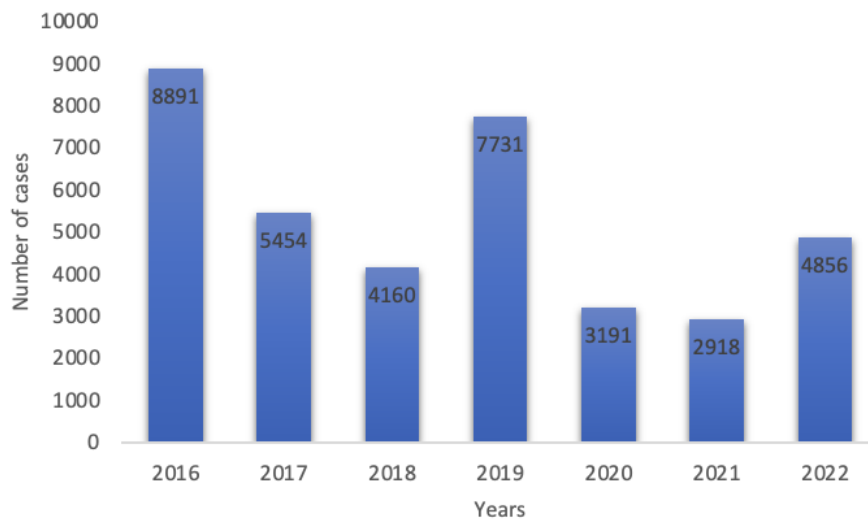
Abstract. This research aims to develop a Geographic Information System (GIS) using the Partitioning Around Medoids (PAM) method in mapping dengue hemorrhagic fever (DHF) cases in North Sumatra Province. DHF has symptoms of high fever, bleeding, and a high mortality rate if not treated quickly. Therefore, mapping DHF cases is very important in efforts to tackle and prevent this disease. The PAM method was used in this study to cluster the DHF case data based on similar characteristics. The first cluster consists of 18 districts/municipalities is a low cluster, the second cluster consists of 3 districts/municipalities is a high cluster, and the third cluster consists of 3 districts/municipalities is a medium cluster. The implementation of PAM is done by using the Minkowski distance calculation method where the application of the distance will be tested with the Silhouette Index on several numbers of clusters. The best number of clusters for PAM implementation is 3 clusters with a Silhouette Index value of 0.5275.

Keywords: DHF, Partitioning Around Medoids, Minkowski.

PENDAHULUAN

DBD atau demam berdarah dengue adalah suatu penyakit demam yang terjadi secara tiba-tiba akibat infeksi oleh virus dengue (Yety Lindawati, dkk, 2021). Penyakit ini dicirikan oleh demam yang berlanjut tinggi selama 2 hingga 7 hari, disertai dengan perdarahan dan kondisi syok, yang jika tidak segera diobati dapat berujung pada kematian (Yety Lindawati, dkk, 2021).

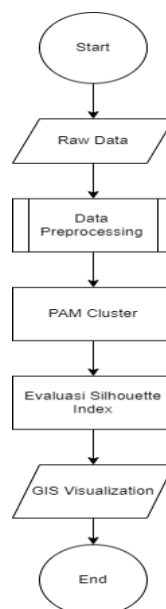
Menurut World Health Organization, setiap tahunnya sekitar 500.000 pasien DBD memerlukan perawatan di rumah sakit, dengan mayoritas dari mereka adalah anak-anak. Diperkirakan sekitar 25% dari pasien anak ini mengalami kematian.



GAMBAR 1. Grafik Jumlah Kasus DBD di Sumatera Utara

Metode

Pada bagian ini, kami akan menguraikan struktur penelitian. Prosesnya digambarkan dalam gambar 2. Dimulai dengan mengumpulkan data, kemudian data tersebut diproses pra-pemrosesan. Data tersebut akan digunakan untuk melakukan proses pengelompokan menggunakan metode PAM. Hasil pengelompokan akan dievaluasi menggunakan metode Silhouette index dan divisualisasikan pada sebuah peta.



GAMBAR 2. Flowchar Penelitian

Pengumpulan Data

Data dikumpulkan dari Dinas Kesehatan Provinsi Sumatera Utara serta beberapa sumber data pelengkap yang berasal dari situs web resmi pemerintah Sumatera Utara seperti Badan Pusat

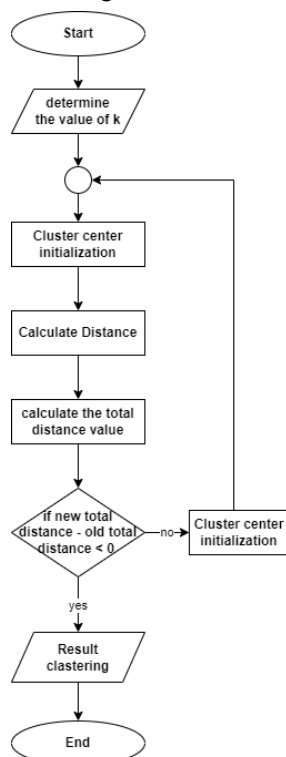
Statistik dan beberapa situs web resmi lainnya. Penelitian ini memanfaatkan data sekunder yang diperoleh dari Dinas Kesehatan Provinsi Sumatera Utara, yang mencakup periode dari Januari 2021 hingga April 2023.

Data Preprocessing

Data Preprocessing adalah tahapan pembersihan dan persiapan data sebelum dilakukan proses analisis data. Dengan melakukan data preprocessing yang baik, hasil analisis data mining akan lebih akurat dan berguna karena ketidakakuratan dan kecacatan data dapat dikurangi (Han & Kamber, 2006). Salah satu langkah dalam pra-pemrosesan data adalah normalisasi data. Normalisasi adalah proses mengubah data mentah sehingga memiliki skala yang seragam atau seimbang untuk setiap fiturnya. Tujuan normalisasi adalah agar setiap fitur memberikan kontribusi yang setara dalam proses pembelajaran algoritma pembelajaran mesin. Dalam normalisasi, ada dua isu utama yang ingin diatasi pada data, yaitu fitur yang dominan yang dapat mendominasi kontribusi dalam model dan adanya outlier yang dapat mempengaruhi hasil secara tidak proporsional (Singh & Singh, 2020). Pada penelitian ini metode normalisasi yang akan digunakan adalah Z-score normalization.

Algoritma PAM

Tahapan ini adalah tahapan membuat klusterisasi dengan Algoritma PAM. Tahapan ini memilih kabupaten atau kota sebagai objeknya adalah kabupaten atau kota dan atributnya yaitu Incidence Rate (IR), Case Fatality Rate (CFR) atau tingkat kematian, Angka Bebas Jentik (ABJ).



GAMBAR 3. PAM Flowchart

Han dan Kamber (2006) menyatakan bahwa tahapan algoritma PAM sebagai berikut (Han & Kamber, 2006).

1. Menentukan jumlah cluster yang akan dibentuk (nilai k).
2. Menginisialisasi pusat cluster awal (medoids) dari masing-masing cluster secara acak.

3. Menghitung setiap data observasi (objek) ke cluster terdekat menggunakan persamaan ukuran jarak Minkowski Distance.
4. Menghitung secara acak objek pada masing-masing cluster sebagai kandidat medoid baru.
5. Menghitung jarak setiap objek yang berada pada masing-masing cluster dengan kandidat medoid baru.
6. Menghitung total simpangan (S) dengan menghitung nilai total distance baru –total distance lama. Jika $S < 0$, maka tukar objek dengan data cluster untuk membentuk sekumpulan k objek baru sebagai medoid.
7. Mengulangi langkah 4 sampai 6 hingga tidak terjadi perubahan medoid, sehingga didapatkan cluster beserta anggota cluster masing-masing.

Peneliti akan melakukan percobaan pada jumlah k dengan mengevaluasi hasil cluster pada nilai k 2,3,4,5,6. Adapun Evaluasi akan dilakukan dengan Silhouette Indeks.

Evaluasi Silhouette Indeks

Silhouette Index (SI) digunakan untuk mengevaluasi data cluster tunggal atau semua cluster. Metode ini juga sering digunakan untuk memvalidasi cluster dengan menggabungkan nilai kohesi dan separasi (Prasetyo, 2014). Adapun persamaan SI adalah sebagai berikut (Justice Panggabean, dkk, 2023).

$$SI = \frac{1}{n} \sum_{i=1}^n S_{(i)} \quad (1)$$

dengan

$$S_{(i)} = \frac{b_{(i)} - a_{(i)}}{\max\{a_{(i)}, b_{(i)}\}} \quad (2)$$

$$b_{(i)} = \min d_{(i,c)} \quad (3)$$

dan

$$a_{(i)} = \frac{1}{|A|-1} \sum_{j \in A, j \neq i} d_{(i,j)} \quad (4)$$

dimana $b(i)$ adalah nilai minimum dari jarak rata-rata objek i dengan semua objek pada cluster lain C, $a(i)$ adalah rata-rata jarak objek ke-i dengan semua objek yang berada di dalam satu cluster A.

Kriteria subjektif pengukuran baik atau tidaknya pengelompokan berdasarkan SI disajikan dalam tabel berikut (Justice Panggabean, dkk, 2023), (Hidayati, dkk, 2021).

TABEL 1. Silhouette Index	
Nilai Silhouette Index (SI)	Struktur
$0.7 < \text{nilai SI} \leq 1$	Struktur Kuat
$0.5 < \text{nilai SI} \leq 0.7$	Struktur Sedang
$0.25 < \text{nilai SI} \leq 0.5$	Struktur Lemah
Nilai SI ≤ 0.25	Tidak Terstruktur

Visualisasi Sistem Informasi Geografis

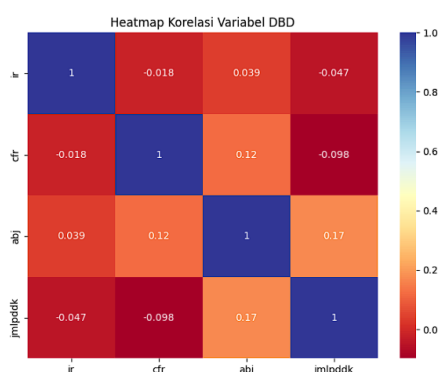
Pada tahapan ini penelitian akan melakukan visualisasi hasil peta dengan menggunakan leaflet. Leafletjs adalah sekelompok JavaScript open source yang dapat digunakan untuk

menyederhanakan pembuatan peta di situs web. Ini merupakan teknologi open source, yang berarti kode-kodenya dapat diakses dan diperiksa untuk memahami cara kerjanya, dan dapat digunakan oleh semua pengguna. Selain itu, siapa pun dapat berpartisipasi dalam proyek ini dengan berkontribusi untuk memperbaiki kode-kodenya (Abdillah, dkk, 2021).

HASIL DAN PEMBAHASAN

Analisis Data

Selanjutnya peneliti melakukan Analisa statistic yaitu menggunakan pearson correlation. Pearson correlation merupakan metode yang paling umum digunakan untuk menghitung korelasi antar variabel. Pada scikit-learn, Pearson correlation dilakukan dengan memanggil fungsi corr(). Dan hasilnya akan berbentuk matriks korelasi seperti berikut.



GAMBAR 4. Heatmap Correlation of Variable DHF

Dari grafik tersebut, terlihat bahwa variabel Jumlah Penduduk dan Incident Rate memiliki korelasi negatif sebesar -0.047, sementara variabel Case Fatality Rate juga memiliki korelasi negatif sebesar -0.098 dengan Incident Rate. Angka-angka ini menunjukkan bahwa ada hubungan negatif antara Jumlah Penduduk dan Incident Rate serta antara Incident Rate dan Case Fatality Rate. Namun, penting dicatat bahwa pengaruhnya dianggap kurang signifikan karena angka korelasi tersebut tergolong rendah. Dengan kata lain, meskipun ada hubungan negatif, dampaknya tidak begitu kuat berdasarkan data yang dianalisis.

Hasil Normalisasi Data

Dalam penelitian ini, data yang dimanfaatkan terdiri dari IR, CFR, dan ABJ pada tiap-tiap kabupaten dan kota di Sumatera Utara, seperti yang terlihat dalam tabel 1. Teknik normalisasi yang dipilih adalah Z-Score. Z-Score normalization merupakan salah satu metode statistika yang digunakan dalam analisis big data. Z-Score, yang juga disebut sebagai nilai standar atau nilai baku, umumnya digunakan dalam data mining untuk mengidentifikasi data yang berada di luar batas normal. Dalam Z-Score, data diubah melalui pembentukan kisaran nilai baru berdasarkan kisaran nilai yang ada dalam dataset sebelumnya. Nilai baru dalam Z-Score dihitung dengan membandingkan nilai data dengan rata-rata dan deviasi standar (Whendasmoro & Joseph, 2022).

TABEL 2. Data Hasil Proses Normalisasi Z-Score

No	Kab/Kota	IR	CRF	ABJ
1.	Medan	0.43	-0.20	1.37

2	Pematang Siantar	1.94	2.250	1.30
3	Binjai	0.58	-0.47	1.36
4	Tanjung Balai	0.38	-0.47	1.33
5	Tebing Tinggi	0.67	0.046	1.28
6	Sibolga	0.86	-0.47	-0.7
7	Padang Sidempuan	-0.5	-0.00	-0.7
8	Deli Serdang	0.19	-0.42	-0.7
9	Langkat	-0.2	-0.47	-0.7
10	Karo	-0.5	-0.47	-0.7
11	Simalungun	0.15	-0.07	-0.7
12	Asahan	-0.4	-0.47	1.11
13	Labuhan Batu	-0.2	-0.47	1.27
14	Tapanuli Utara	-0.6	-0.47	1.29
15	Tapanuli Tengah	-0.4	-0.29	1.38
16	Tapanuli Selatan	-0.8	4.098	-0.7
17	Nias	0.31	-0.47	-0.7
18	Dairi	-0.5	-0.47	-0.7
19	Toba	0.03	1.468	0.49
20	Mandailing Natal	-0.8	1.811	1.30
21	Nias Selatan	-0.6	-0.47	-0.7
22	Pak-Pak Bharat	1.10	-0.21	-0.7
23	Humbahas	-0.8	-0.47	-0.7
24	Samosir	-0.4	-0.47	-0.7
25	Serdang Bedagai	-0.7	-0.47	1.37
26	Batubara	-0.1	0.192	1.06
27	Padang Lawas	-0.3	-0.47	-0.7
28	Padang Lawas Utara	-0.6	-0.47	-0.7
29	Labuhan Batu Selatan	-0.7	-0.47	-0.7

30	Labuhan Batu Utara	-0.5	-0.47	-0.7
31	Nias Utara	-0.1	-0.47	-0.7
32	Nias Barat	-0.5	0.869	-0.7
33	Gunungsitoli	4.37	-0.47	-0.7

Partitioning Around Medoid (PAM)

Pada penelitian ini, proses data mining yang dilakukan adalah clustering dengan metode PAM. yang mana tujuannya adalah untuk mendapatkan pemetaan objek yang memiliki nilai dan karakteristik data yang sama. Pada implemetasi metode PAM peneliti akan melakukan pengujian dengan menggunakan 2, 3, 4, 5 jumlah cluster dan pada percobaan peneliti memilih Minkowski Distance sebagai metode perhitungan jarak.

Untuk melakukan evaluasi peneliti menggunakan metode Silhouette Index. Hasil cluster yang dievaluasi adalah berdasarkan beberapa jumlah cluster yaitu 2, 3, 4, 5. Adapun hasil pengujian Silhouette Index sebagai berikut.

TABEL 3. Hasil Pengujian Silhouette Index

n_cluster	Nilai Silhouette Index
2	0.4699
3	0.5275
4	0.4360
5	0.3138

Dari table 3 ditampilkan nilai SI yang didapat dari pengujian jumlah cluster. Jika dilihat dari table n_cluster 3 mendapatkan nilai SI tertinggi dengan nilai 0.5275. Dikarenakan hal tersebut jumlah n_cluster 3 adalah yang terbaik dan akan digunakan untuk mengclusterkan data DBD dengan PAM yang selanjutnya divisualisasikan dalam bentuk peta warna. Adapun hasil cluster sebagai berikut.

TABEL 4. Hasil Clustering Berdasarkan Minkwoski Distance

No	Kab/Kota	CI	C2	C3	Cluster
1.	Medan	2.21	2.62	0.70	3
2	Pematang Siantar	3.49	0	3.13	2
3	Binjai	2.23	2.83	0.83	3
4	Tanjung Balai	2.16	2.88	0.63	3
5	Tebing Tinggi	2.18	2.33	0.97	3

6	Sibolga	1.27	3.13	2.17	1
7	Padang Sidempuan	0.47	3.30	2.07	1
8	Deli Serdang	0.60	3.23	2.07	1
9	Langkat	0.14	3.42	2.06	1
10	Karo	0.14	3.56	2.07	1
11	Simalungun	0.62	3.02	2.07	1
12	Asahan	1.91	3.25	0.24	3
13	Labuhan Batu	2.06	3.13	0	3
14	Tapanuli Utara	2.09	3.37	0.43	3
15	Tapanuli Tengah	2.17	3.11	0.25	3
16	Tapanuli Selatan	4.57	3.33	4.71	2
17	Nias	0.71	3.23	2.08	1
18	Dairi	0.19	3.58	2.07	1
19	Toba	2.12	1.99	1.98	3
20	Mandailing Natal	2.77	2.76	2.29	3
21	Nias Selatan	0.20	3.59	2.07	1
22	Pak-Pak Bharat	1.51	2.91	2.24	1
23	Humbahas	0.44	3.72	2.08	1
24	Samosir	0	3.49	2.06	1
25	Serdang Bedagai	2.16	3.38	0.45	3
26	Batubara	1.89	2.60	0.67	3
27	Padang Lawas	0.08	3.45	2.06	1
28	Padang Lawas Utara	0.25	3.61	2.07	1
29	Labuhan Batu Selatan	0.38	3.68	2.08	1
30	Labuhan Batu Utara	0.10	3.54	2.07	1
31	Nias Utara	0.29	3.36	2.06	1
32	Nias Barat	1.34	2.98	2.24	1
33	Gunungsitoli	4.77	3.52	4.75	2

Dengan penggambaran peta sebaran sebagai berikut.

DAFTAR PUSTAKA

- Abdillah, M. Z., Nawangnugraeni, D. A., Hakim, A., & Yuniarto, P. (2021). Geographic Information System (GIS) For Mapping Greenpark Using Leaflet Js. *Jurnal Teknik Informatika Kaputama (JTIK)*, 5(2).
- Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques. 2nd Edition* (3rd ed.). Morgan Kaufmann Publishers.
- Hidayati, R., Zubair, A., Hidayat Pratama, A., & Indana, L. (2021). Analisis Silhouette Coefficient pada 6 Perhitungan Jarak K-Means Clustering Silhouette Coefficient Analysis in 6 Measuring Distances of K-Means Clustering (Vol. 20, Issue 2).
- Justice Panggabean, M., Molliq Rangkuti, Y., & Muslim Karo Karo, I. (2023). Design of a geographic information system (gis) for the spread of covid-19 disease in medan city. In *Jurnal Mantik* (Vol. 6, Issue 4). Online.
- Marlina, D., Fauzer Putri, N., Fernando, A., & Ramadhan, A. (2018). Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak. *Jurnal CoreIT*, 4(2), 64–71.
- Masnur, Alam, S., & Ihsar, M. (2022). Aplikasi Sistem Informasi Geografis (SIG) Pemetaan Lahan Pertanian Dan Komoditas Hasil Panen Di Kabupaten Sidrap Berbasis Web. *Jurnal Sintaks Logika*, 2(1), 229–235.
- Meiriza, A., Lestari, E., Putra, P., Oktadini, N. R., & Audya, M. (2022). Pemanfaatan Data Mining Dalam Penentuan Penyuluhan Penyakit Stunting Menggunakan Partitioning Around Medoids (PAM). *Teika*, 12(2), 89–96.
- Prasetyo, E. (2014). Mengolah Data Menjadi Informasi Menggunakan Matlab. ANDI.
- Singh, D., & Singh, B. (2020). Investigating the impact of data normalization on classification performance. *Applied Soft Computing*, 97. <https://doi.org/10.1016/j.asoc.2019.105524>
- Whendasromo, R. G., & Joseph, J. (2022). Analisis Penerapan Normalisasi Data Dengan Menggunakan Z-Score Pada Kinerja Algoritma K-NN. *JURIKOM (Jurnal Riset Komputer)*, 9(4), 872. <https://doi.org/10.30865/jurikom.v9i4.4526>
- Wulandari, S., & Dwitiyanti, N. (2020). Implementasi Algoritma Clustering Partitioning Around Medoid (PAM) Dalam Clustering Virus Mers-Cov. www.ncbi.nlm.nih.gov.
- Yety Lindawati, N., Murtisiwi, L., Aisyah Rahmania, T., Nita Damayanti, P., Marentina Widyasari, F., Studi, P. S., Tinggi Ilmu Kesehatan Nasional, S., & Tengah, J. (2021). Upaya Peningkatan Pengetahuan Masyarakat Dalam Rangka Pencegahan dan Penanggulangan DBD Di Desa Dlingo, Mojosongo, Boyolali. 4.