

ANALISIS CLUSTER WILAYAH RAWAN BENCANA ALAM DI JAWA BARAT MENGUNAKAN ALGORITMA K-MEDOIDS

Ocha Meilyla

Program Studi S1 Matematika, FMIPA, Universitas Negeri Surabaya

e-mail: ochameilyla.21041@mhs.unesa.ac.id

A'yunin Sofro

Program Studi S1 Matematika, FMIPA, Universitas Negeri Surabaya

e-mail: ayuninsofro@unesa.ac.id*

Abstrak

Bencana alam merupakan peristiwa yang dapat menimbulkan dampak signifikan terhadap kehidupan masyarakat, baik dari segi kerugian ekonomi, sosial, maupun infrastruktur. Oleh karena itu, identifikasi wilayah rawan bencana menjadi langkah krusial dalam upaya mitigasi dan perencanaan kebijakan penanggulangan bencana. Penelitian ini bertujuan untuk melakukan analisis cluster wilayah rawan bencana alam di Jawa Barat menggunakan algoritma *K-Medoids*, yang dikenal lebih robust terhadap outlier dibandingkan metode clustering lainnya. Dalam penelitian ini, pengelompokan dilakukan berdasarkan beberapa variabel yang berpengaruh dalam menentukan tingkat rawan bencana alam di suatu wilayah. Untuk meningkatkan keakuratan hasil clustering, perhitungan jarak dilakukan menggunakan dua metrik, yaitu *Euclidean distance* dan *Manhattan distance*. Hasil dari proses *clustering* kemudian divalidasi menggunakan *Davies-Bouldin index* guna menentukan kualitas dan optimalitas pengelompokan yang diperoleh. Hasil penelitian menunjukkan bahwa metode *K-Medoids* mampu mengelompokkan wilayah rawan bencana secara efektif dengan pola yang lebih jelas. Diperoleh 2 *cluster* yang menggambarkan tingkat kerawanan bencana di berbagai daerah di Jawa Barat. *Cluster 1* menunjukkan wilayah dengan tingkat rawan bencana yang tinggi dan *cluster 2* menunjukkan wilayah dengan tingkat rawan bencana yang rendah. Hasil perhitungan *Davies-Bouldin index* menunjukkan bahwa penggunaan *Manhattan distance* lebih optimal dibandingkan dengan *Euclidean distance* dengan nilai DBI sebesar 0,299043. Temuan ini dapat menjadi referensi dalam perumusan kebijakan mitigasi bencana yang lebih tepat sasaran serta meningkatkan kesiapsiagaan masyarakat dalam menghadapi potensi bencana di masa mendatang.

Kata Kunci: Analisis Cluster, Wilayah Rawan Bencana, *K-Medoids*, *Davies-Bouldin index*.

Abstract

Natural disasters are events that can have a significant impact on people's lives, both in terms of economic, social, and infrastructure losses. Therefore, identifying disaster-prone areas is a crucial step in disaster mitigation and policy planning efforts. This study aims to conduct a cluster analysis of disaster-prone areas in West Java using the *K-Medoids* algorithm, which is known to be more robust to outliers than other clustering methods. In this study, grouping is carried out based on several variables that influence the level of disaster risk in an area. To improve the accuracy of the clustering results, distance calculations are carried out using two metrics, namely *Euclidean Distance* and *Manhattan Distance*. The results of the clustering process are then validated using the *Davies-Bouldin Index* to determine the quality and optimality of the grouping obtained. The results show that the *K-Medoids* method is able to group disaster-prone areas effectively with a clearer pattern. 2 clusters were obtained that describe the level of disaster risk in various regions in West Java. Cluster 1 shows areas with a high level of disaster risk and cluster 2 shows areas with a low level of disaster risk. The *Davies-Bouldin Index* calculation results show that the use of *Manhattan Distance* is more optimal compared to *Euclidean Distance* with a DBI value of 0.299043. This finding can be a reference in formulating more targeted disaster mitigation policies and increasing community preparedness in facing potential disasters in the future.

Keywords: Cluster Analysis, Disaster Prone Areas, *K-Medoids*, *Davies-Bouldin index*.

PENDAHULUAN

Bencana alam merupakan fenomena yang dapat menyebabkan kerusakan infrastruktur, hilangnya nyawa, serta gangguan ekonomi dan sosial yang signifikan. Indonesia, sebagai negara kepulauan yang

berada di jalur Cincin Api Pasifik, memiliki tingkat risiko bencana alam yang tinggi. Berdasarkan data yang dicatat oleh Badan Nasional Penanggulangan Bencana (BNPB), sepanjang tahun 2024 telah terjadi sebanyak 203 kejadian bencana alam di wilayah Jawa Barat. Wilayah Jawa Barat sering mengalami berbagai

jenis bencana alam seperti banjir, longsor, gempa bumi, dan kekeringan, yang berdampak pada kehidupan masyarakat dan pembangunan daerah. Oleh karena itu, pemetaan daerah rawan bencana menjadi sangat penting sebagai dasar dalam perencanaan kebijakan mitigasi dan tanggap darurat. Mitigasi merupakan upaya yang dilakukan untuk menekan risiko bencana serta meminimalkan dampak negatifnya terhadap masyarakat, baik dalam hal korban jiwa maupun kerugian materi, yang dapat memengaruhi kehidupan dan aktivitas manusia (Sutiyo dan Eviany, 2023). Salah satu pendekatan yang dapat digunakan untuk memahami pola kerawanan bencana alam adalah dengan menggunakan metode analisis cluster.

Melalui metode analisis *cluster*, pembagian wilayah berdasarkan pola kerawanan dapat dilakukan dengan mempertimbangkan berbagai variabel penting yang berpengaruh dalam menentukan tingkat rawan bencana alam di suatu daerah. Terdapat penelitian (Murdiaty et al., 2020) yang menggunakan algoritma *K-Means* sebagai pengelompokan data bencana alam berdasarkan wilayah, waktu, jumlah korban dan kerusakan fasilitas. Penelitian selanjutnya (Maulana Ichsan et al., 2024) melakukan pengelompokan data wilayah rawan bencana alam di pulau jawa, lalu penelitian berikutnya (Halik & Septiana, 2022) melakukan analisa data dengan memprediksi daerah rawan bencana alam di jawa barat menggunakan algoritma *K-Means Clustering*. Penelitian tersebut menunjukkan bahwa analisis *cluster* terbukti efektif dalam mengidentifikasi pola yang terdapat dalam data bencana alam. Oleh karena itu penelitian ini akan menerapkan analisis *cluster* dengan inovasi dalam metode yang diterapkan, sehingga memiliki perbedaan dengan penelitian sebelumnya. Tujuan utama dari pendekatan ini adalah untuk memperoleh hasil yang lebih baik dalam proses pengelompokan wilayah berdasarkan tingkat rawan bencana alam.

K-Means adalah algoritma pengelompokan yang membagi objek ke dalam satu atau lebih kelompok berdasarkan kesamaan karakteristik. Objek dengan karakteristik yang lebih mirip akan dikelompokkan ke dalam cluster yang sama. Sementara itu, *K-Medoids* merupakan varian pengembangan dari algoritma *K-Means* yang berfungsi untuk meningkatkan efisiensi yang didasarkan pada pemilihan prototipe representatif yang lebih stabil dalam setiap *cluster*.

Berdasarkan penelitian (Arora et al., 2016) pengelompokan data dalam skala besar menggunakan *K-Medoids* dan *K-Means* menunjukkan bahwa *K-Medoids* memiliki keunggulan dalam waktu eksekusi yang lebih efisien serta lebih tahan terhadap outlier. Selanjutnya penelitian (Karo Karo et al., 2023) menunjukkan bahwa penggunaan algoritma *K-Medoids* lebih unggul dibandingkan *K-Means* dalam proses pengelompokan. Keunggulan tersebut menjadi salah satu karakteristik utama dari algoritma *K-Medoids*. Dalam proses pengelompokannya, *K-Medoids* memerlukan perhitungan jarak yang digunakan dalam menentukan *cluster*. Pemilihan metrik jarak yang tepat sangat penting, karena dapat mempengaruhi kualitas hasil akhir dari proses clustering. Berdasarkan penelitian (Sunge et al., 2020) proses pengelompokan menggunakan *K-Medoids*, diperoleh hasil yang optimal ketika menggunakan metrik jarak *Manhattan* dan *Euclidean*.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk menganalisis cluster wilayah rawan bencana alam di Jawa Barat. Analisis ini dilakukan untuk memahami karakteristik wilayah rawan bencana sebagai dasar perencanaan mitigasi bencana yang lebih efektif dan terarah, serta meningkatkan kesiapsiagaan masyarakat. Metode analisis cluster yang digunakan adalah *K-Medoids* dengan perhitungan jarak *Euclidean* dan *Manhattan*. Untuk menentukan hasil *cluster* yang paling optimal, hasil analisis *K-Medoids* dengan kedua jenis perhitungan jarak tersebut akan divalidasi menggunakan *Davies-Bouldin index*. Hasil optimal ini akan menunjukkan pengelompokan wilayah berdasarkan tingkat kerawanan bencana alam di Provinsi Jawa Barat.

KAJIAN TEORI

BENCANA ALAM

Bencana alam merupakan peristiwa yang seringkali sulit diprediksi waktu kejadiannya dan dapat menimbulkan dampak yang sangat merugikan. Peristiwa ini menyebabkan banyak korban jiwa serta kerusakan fisik pada infrastruktur (Jefri Ardian Nugroho et al., 2010). Dampaknya tidak hanya bersifat jangka pendek, tetapi juga berlanjut dalam bentuk penurunan kualitas hidup masyarakat. Beberapa wilayah memiliki tingkat kerawanan yang lebih tinggi terhadap bencana alam akibat faktor

geografis, kondisi lingkungan dan kesiapan infrastruktur. Untuk mengidentifikasi daerah rawan bencana, diperlukan analisis berbasis data menggunakan beberapa indikator utama, seperti jumlah bencana alam yang terjadi di suatu wilayah sebagai ukuran frekuensi kejadian, jumlah korban yang terdampak sebagai indikator tingkat risiko terhadap masyarakat, jumlah kerusakan rumah sebagai gambaran dampak terhadap permukiman, serta jumlah kerusakan fasilitas yang menunjukkan seberapa besar gangguan terhadap infrastruktur publik dan layanan dasar. Dengan menganalisis variabel-variabel ini, pemetaan daerah rawan bencana dapat dilakukan secara lebih akurat, sehingga membantu dalam perencanaan strategi mitigasi yang lebih efektif. Langkah-langkah seperti penguatan infrastruktur, penyusunan kebijakan berbasis risiko, serta peningkatan kesiapsiagaan masyarakat dapat diimplementasikan untuk mengurangi dampak bencana dan meningkatkan ketahanan wilayah terhadap bencana alam di masa depan. Untuk mengelompokkan daerah berdasarkan tingkat kerawannya, diperlukan metode analisis data yang tepat. Salah satu pendekatan yang dapat digunakan adalah analisis *cluster*, yang memungkinkan pengelompokan wilayah dengan karakteristik serupa berdasarkan indikator yang telah ditentukan. Tahap awal dalam analisis ini adalah menentukan jumlah *cluster* yang optimal menggunakan *silhouette score*, sehingga hasil pengelompokan dapat lebih akurat dan representatif.

SILHOUETTE SCORE

Penentuan jumlah *cluster* yang optimal merupakan langkah krusial yang mempengaruhi validitas hasil pengelompokan dan pengambilan keputusan berbasis data. Salah satu metode yang umum digunakan untuk mengevaluasi kualitas clustering adalah metode *silhouette score*. Metode ini mengukur tingkat kohesi (cohesion) dalam satu *cluster* dan tingkat pemisahan (separation) antar *cluster*. Kohesi dihitung berdasarkan kedekatan objek-objek dalam satu *cluster*, sedangkan pemisahan ditentukan dari rata-rata jarak suatu objek terhadap *cluster* terdekatnya. Dengan menggunakan metode *Silhouette Score*, jumlah *cluster* yang paling optimal dapat ditentukan berdasarkan nilai evaluasi yang

diperoleh. Berikut kriteria pengukuran *Silhouette Score* (Kaufman & Rousseeuw, 1990):

Tabel 1. Kriteria *Silhouette Score*

Nilai koefisien Silhouette	Kriteria Cluster
0,71 – 1,00	Strong
0,51 – 0,70	Good
0,26 – 0,50	Weak
0,00 – 0,25	Bad

ANALISIS CLUSTER

Analisis *cluster* adalah metode yang digunakan untuk mengelompokkan data atau objek berdasarkan pada kesamaan karakteristik. Setiap kelompok berisi data atau objek yang mirip satu sama lain, sementara disaat yang sama kelompok tersebut berbeda dengan kelompok yang lain dengan aspek yang lain. Analisis *cluster* membantu menyederhanakan dan memahami struktur data secara lebih mendalam (Everitt & Leese, 2011). Pada analisis *cluster* terdapat dua metode utama, yaitu metode hirarki dan non-hirarki. Metode hirarki adalah pengelompokan data dengan berdasarkan kemiripan data atau objek namun penentuan jumlah kelompok tidak ditentukan sebelumnya. Metode non-hirarki adalah teknik pengelompokan data dengan berdasarkan kemiripan data atau objek namun untuk penentuan jumlah kelompok sudah ditentukan sebelumnya menggunakan metode, seperti *silhouette score*, *elbow*, dll. Hal ini menjadi keunggulan dari metode non-hirarki karena penentuan jumlah *cluster*nya lebih jelas dan proses pengelompokan lebih cepat dibandingkan dengan metode hirarki. Salah satu contoh metode non-hirarki adalah *K-Medoids*. Berdasarkan penelitian menunjukkan bahwa menggunakan algoritma *K-Medoids* lebih baik dalam mengelompokkan data (Karo Karo et al., 2023). Berikut penjelasan mengenai metode *K-Medoids*.

K-MEDOIDS

K-Medoids adalah salah satu algoritma yang dirancang untuk meminimalkan fungsi objektif yang sudah ditentukan sebelumnya. Fungsi objektif ini diukur dengan berdasarkan total jarak antara titik-titik data pada suatu *cluster* dengan *medoid* (pusat *cluster*). *K-Medoids* menggunakan titik data aktual sebagai prototipe, hal ini memberikan keunggulan pada *K-Medoids* dalam menangani outlier pada data.

Proses Algoritma *K-Medoids* berjalan secara iteratif, dengan setiap iterasi algoritma akan memilih medoid, lalu setiap titik data diterapkan ke *medoid* terdekat berdasarkan perhitungan jarak tertentu. Setelah *cluster* terbentuk, maka *medoid* akan diperbarui disetiap *cluster* dengan titik data lain yang dapat meminimalkan total jarak pada *cluster* (Aggarwal & K.Reddy, 2014). Berikut langkah-langkah dari algoritma *K-Medoids* :

1. Menentukan nilai *k* atau jumlah *cluster* yang akan digunakan dengan *silhouette score*.
2. Menentukan pusat *cluster* (*medoid*) awal secara acak dari setiap *cluster* berdasarkan data aktual pada data yang akan digunakan.
3. Mengalokasikan semua objek ke *cluster* menggunakan perhitungan jarak, seperti menggunakan jarak *euclidean* atau *manhattan*.
4. Memilih objek *non-medoid* secara acak untuk setiap *cluster* sebagai kandidat *medoid* baru.
5. Menghitung total jarak untuk setiap objek pada masing-masing *cluster*.
6. Menghitung *S* (total simpangan) dengan cara menghitung total jarak baru dikurangi dengan total jarak lama.
7. Apabila $S < 0$, maka tukar *medoid* dengan x_i untuk membentuk set objek representatif yang baru, apabila $S > 0$ maka iterasi akan dihentikan dan hasil clustering pada iterasi terakhir dimana total jarak tercapai sebelum $S > 0$ terdeteksi.
8. Ulangi proses pada langkah ke-3 hingga ke-7, sampai tidak ada perubahan *medoid* dan diperoleh anggota *cluster*.

Metode *K-Medoids* berperan dalam menentukan representasi klaster dengan meminimalkan fungsi objektif. Dalam analisis menggunakan *K-Medoids*, perhitungan jarak yang digunakan mencakup *Euclidean distance* dan *Manhattan distance*. Berikut adalah penjelasan lebih rinci mengenai perhitungan jarak tersebut

EUCLIDEAN DISTANCE

Euclidean distance adalah metode yang populer digunakan untuk mengukur seberapa jauh dua titik terpisah dalam ruang, seperti bidang datar atau dalam ruang tiga dimensi. Karena kesederhanaannya *Euclidean distance* menjadi pilihan utama dalam banyak algoritma dan analisis data (Nishom, 2019). Berikut adalah persamaan jarak euclidean untuk mengukur kemiripan data (Grabusts, 2011):

$$d_{ij} = \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2}$$

Dimana :

d_{ij} = jarak antara objek *i* dan *j*

x_{ik} = nilai objek *i* dalam variabel ke-*k*

x_{jk} = nilai objek *j* dalam variabel ke-*k*

m = jumlah data

MANHATTAN DISTANCE

Manhattan distance adalah salah satu metode perhitungan jarak yang digunakan untuk mengukur perbedaan absolut antara koordinat sepasang objek pada suatu ruang. Metode ini menjumlahkan perbedaan pada setiap dimensi secara terpisah, menghasilkan nilai yang merepresentasikan jarak yang ditempuh sepanjang sumbu koordinat. Sifat ini menjadikan *manhattan distance* relevan dalam konteks di mana pergerakan terbatas pada arah orthogonal (Nishom, 2019). Berikut adalah persamaan jarak *manhattan* untuk mengukur kemiripan data (Grabusts, 2011) :

$$d_{ij} = \sum_{k=1}^m |x_{ik} - x_{jk}|$$

Dimana :

d_{ij} = jarak antara objek *i* dan *j*

x_{ik} = nilai objek *i* dalam variabel ke-*k*

x_{jk} = nilai objek *j* dalam variabel ke-*k*

m = jumlah data

Setelah dilakukan perhitungan jarak, maka untuk menentukan hasil cluster yang paling optimal akan menggunakan *Davies-Bouldin Index*. Berikut merupakan penjelasan mengenai *Davies-Bouldin Index* lebih lanjut.

DAVIES-BOULDIN INDEX

Davies-Bouldin Index (DBI) merupakan salah satu metode evaluasi yang berfungsi untuk mengukur kualitas dari hasil *clustering* pada data. Pertama kali metode ini diperkenalkan oleh David L. Davies dan Donald W. Bouldin pada tahun 1979, sejak saat itu metode ini menjadi salah satu teknik evaluasi *clustering* yang populer digunakan hingga saat ini. Perhitungan *Davies-Bouldin Index* (DBI) dilakukan dengan membandingkan *intra-cluster* dan *inter-cluster* dengan menggunakan rasio *inter-cluster distance* dan *intra-cluster distance* (Meiriza et al., 2023). Hasil

clustering terbaik ketika *inter-cluster distance* bernilai tinggi dan *intra-cluster distance* bernilai rendah. Oleh karena itu, nilai dari *Davies-Bouldin Index* (DBI) yang lebih kecil mengindikasikan bahwa kualitas hasil *clustering* yang lebih baik (Kartikasari, 2021). Berikut langkah-langkah dari perhitungan *Davies-Bouldin Index* (DBI) (Muthmainah et al., 2024):

1. Mengukur jarak *intra-cluster* dengan menghitung antara setiap data dalam *cluster* terhadap titik pusatnya.

$$SSW_i = \frac{1}{n} \sum_{j=1}^n d(x_j, c_i)$$

dimana n adalah banyaknya data pada *cluster* i . x_j adalah data pada *cluster* i . Dan c_i adalah pusat dari *cluster* i .

2. Menentukan jarak *inter-cluster* dengan menghitung jarak antar *cluster*.

$$SSB_{i,j} = d(c_i, c_j)$$

dimana c_i merupakan pusat pada *cluster* i dan c_j merupakan pusat pada *cluster* j , dengan $i \neq j$

3. Pada setiap pasangan *cluster*, sebaran clusternya ditambahkan lalu totalnya akan dibagi dengan jarak antara kedua *cluster*.

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}}$$

rumus ini memberikan ukuran seberapa "dekat" dua *cluster* tersebut dalam konteks sebaran data. Jika nilai dari R_{ij} rendah, maka hal ini menunjukkan bahwa kedua *cluster* tersebut serupa dalam hal sebaran, yang berarti *clustering* yang dihasilkan efektif.

4. Perhitungan ukuran kesamaan *cluster* maksimum

$$R_i = \max_{i \neq j} R_{ij}$$

5. Rumus DBI sebagai rata-rata ukuran kesamaan *cluster* maksimum.

$$DBI = \frac{1}{K} \sum_{i=1}^k R_i$$

dimana k merupakan banyaknya *cluster*.

METODE

DATA PENELITIAN

Penelitian ini dilakukan secara kuantitatif menggunakan data sekunder untuk melakukan analisis *cluster* yang diperoleh dari website Badan

Nasional Penanggulangan Bencana, [Data Online - BNPB](#), yaitu data jumlah kejadian bencana, jumlah korban, jumlah kerusakan rumah, dan jumlah kerusakan fasilitas di wilayah provinsi Jawa Barat yang didasarkan pada data informasi bencana Indonesia pada tahun 2024. Variabel yang digunakan adalah jumlah kejadian bencana, jumlah korban, jumlah kerusakan rumah, dan jumlah kerusakan fasilitas.

RANCANGAN PENELITIAN

Untuk mencapai tujuan penelitian, terdapat rancangan penelitian yang harus dilakukan. Langkah pertama adalah pengumpulan data yang diperoleh dari website resmi Badan Nasional Penanggulangan Bencana (BNPB). Selanjutnya, akan digunakan metode *Silhouette Score* untuk menentukan jumlah *cluster* optimal. Setelah diketahui jumlah *cluster* optimal, maka akan dilakukan analisis *cluster* menggunakan algoritma *K-Medoids* dengan perhitungan jarak *Euclidean* dan *Manhattan*. Hasil analisis *cluster* diketahui dan langkah berikutnya yaitu menentukan kualitas hasil *clustering* menggunakan *Davies-Bouldin Index*. Setelah diketahui nilai dari *Davies-Bouldin Index*, maka akan dibandingkan dan diambil hasil yang paling optimal.

HASIL DAN PEMBAHASAN

SILHOUETTE SCORE

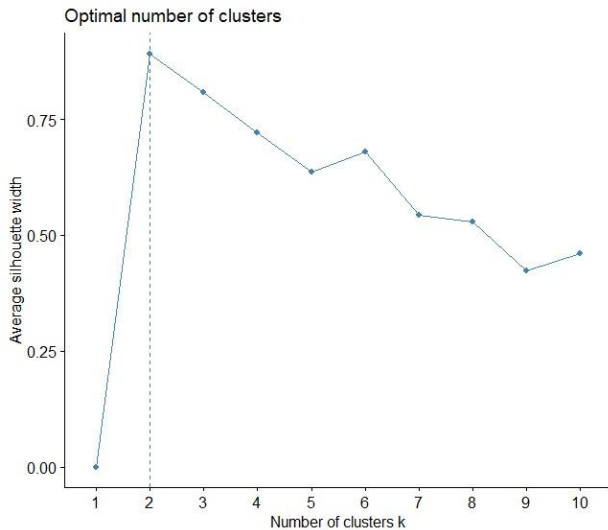
Data bencana alam yang digunakan dalam penelitian ini terdiri dari empat variabel dengan total 26 data yang masing-masing mewakili wilayah di Provinsi Jawa Barat. Variabel yang dianalisis dalam penelitian ini meliputi :

Tabel 2. Simbol dan variabel

Simbol	Nama Variabel
X_1	Jumlah Kejadian bencana alam
X_2	Jumlah Korban
X_3	Jumlah kerusakan rumah
X_4	Jumlah kerusakan fasilitas

Data ini mencerminkan berbagai aspek terkait bencana alam yang terjadi di wilayah tertentu dan digunakan untuk mendukung proses analisis *clustering*. Setiap variabel dipilih berdasarkan relevansinya dalam mengukur dampak dan karakteristik bencana, sehingga dapat memberikan wawasan yang lebih komprehensif terhadap pola

kejadian bencana. Sebelum melakukan proses analisis *clustering*, langkah pertama yang dilakukan adalah menentukan jumlah *cluster* yang optimal. Penentuan jumlah *cluster* ini menggunakan Silhouette Score, yang hasilnya akan ditampilkan pada grafik berikut :



Gambar 1. Grafik Silhouette Score

Berdasarkan hasil grafik perhitungan silhouette score, jumlah cluster yang optimal diperoleh saat $k=2$. Hal ini dikarenakan nilai *Silhouette Score* pada $k=2$ berada di atas 0,75, yang menurut kriteria pengukuran nilai *Silhouette Score* menunjukkan kualitas *clustering strong* yang berarti sangat baik. Dibandingkan dengan jumlah cluster yang lainnya, nilai *Silhouette Score* tertinggi dicapai saat $k=2$, sehingga jumlah cluster ini dianggap paling optimal untuk digunakan dalam proses analisis selanjutnya.

K-MEDOIDS DENGAN EUCLIDEAN DISTANCE

Proses selanjutnya adalah analisis menggunakan algoritma *K-Medoids* dengan perhitungan jarak *Euclidean*. Tahapan algoritma *K-Medoids* mengikuti prosedur yang telah dijelaskan dalam kajian pustaka, dengan jarak *Euclidean* sebagai metrik pengukuran. Hasil clustering yang diperoleh disajikan sebagai berikut :

Table 3. Hasil *Clustering K-Medoids* dengan *Euclidean distance*

Cluster	Jumlah anggota	Nama anggota
1	9	Bogor, Bandung, Cirebon, Majalengka, Indramayu, Karawang, Bekasi, Bandung barat, Kota Bekasi.
2	17	Sukabumi, Cianjur, Garut, Tasikmalaya, Ciamis, Kuningan, Sumedang, Subang, Purwakarta, Pangandaran, Kota Bogor, Kota Sukabumi, Kota Bandung, Kota Cirebon, Kota Cimahi, Kota Tasikmalaya, Kota Banjar.

Berdasarkan hasil clustering menggunakan *Euclidean Distance*, wilayah di Jawa Barat terbagi menjadi dua cluster yang mencerminkan kesamaan karakteristik berdasarkan variabel yang dianalisis. Cluster 1 terdiri dari 9 wilayah, sedangkan Cluster 2 mencakup 17 wilayah. Pembagian ini menunjukkan bahwa wilayah dalam masing-masing cluster memiliki pola yang lebih mirip satu sama lain dalam hal tingkat risiko atau dampak bencana.

K-MEDOIDS DENGAN MANHATTAN DISTANCE

Setelah proses analisis menggunakan algoritma *K-Medoids* dengan perhitungan jarak *Euclidean*, maka akan dilakukan proses analisis yang sama dengan menggunakan perhitungan jarak *Manhattan*. Tahapan algoritma *K-Medoids* mengikuti prosedur yang telah dijelaskan dalam kajian pustaka, dengan jarak *Manhattan* sebagai metrik pengukuran. Hasil *clustering* yang diperoleh disajikan sebagai berikut :

Table 4. Hasil *Clustering K-Medoids* dengan *Manhattan distance*

Cluster	Jumlah anggota	Nama anggota
1	9	Bogor, Bandung, Cirebon, Majalengka, Indramayu, Karawang, Bekasi, Bandung barat, Kota Bekasi.

2	17	Sukabumi, Cianjur, Garut, Tasikmalaya, Ciamis, Kuningan, Sumedang, Subang, Purwakarta, Pangandaran, Kota Bogor, Kota Sukabumi, Kota Bandung, Kota Cirebon, Kota Cimahi, Kota Tasikmalaya, Kota Banjar.
---	----	--

Berdasarkan hasil metode perhitungan jarak *Manhattan Distance*, hasil *clustering* tetap menunjukkan pembagian yang sama seperti pada perhitungan menggunakan jarak *Euclidean*. Hal ini mengindikasikan bahwa pola hubungan antar wilayah cukup stabil, Maka untuk mengetahui perbedaan hasil *clustering* antara penggunaan jarak *Euclidean* dan *Manhattan* akan dilakukan validasi dengan melihat nilai dari *Davies-Bouldin index*.

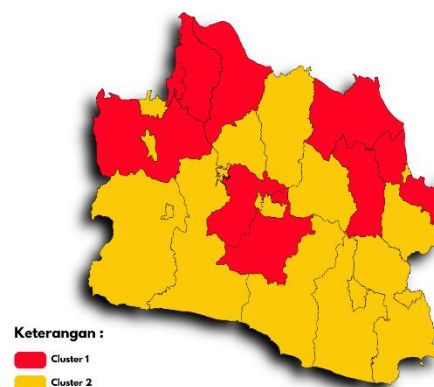
DAVIES-BOULDIN INDEX

Evaluasi hasil *clustering* dalam penelitian ini dilakukan menggunakan *Davies-Bouldin Index*. Dalam penelitian terdapat hasil analisis menggunakan algoritma *K-Medoids* dengan *Euclidean distance* dan *K-Medoids* dengan *Manhattan distance* dengan total 2 *cluster*, yaitu *cluster 1* sebanyak 9 data dan *cluster 2* sebanyak 17 data. Hasil DBI pada *K-Medoids* dengan *Euclidean distance* memiliki nilai 0,305785, sedangkan *K-Medoids* dengan *Manhattan distance* memiliki nilai 0,299043 yang ditunjukkan pada tabel 5.

Table 5. Hasil *Davies-Bouldin Index*

K-Medoid dengan metrik jarak	Nilai DBI
<i>Euclidean distance</i>	0,305785
<i>Manhattan distance</i>	0,299043

Hasil analisis *cluster* dikatakan optimal jika nilai DBI memiliki nilai yang kecil (mendekati 0 dan non negatif), sehingga berdasarkan hasil DBI pada dua jarak tersebut, ditemukan bahwa penggunaan *K-Medoids* dengan *Manhattan distance* lebih baik dari pada penggunaan *K-Medoids* dengan *Euclidean distance*. Sehingga hasil *clustering* yang digunakan seperti pada gambar berikut:



Gambar 2. Peta hasil clustering

Berdasarkan Gambar 2. *Cluster 1* ditandai dengan warna merah yang berarti wilayah tersebut memiliki tingkat rawan bencana alam yang tinggi, yaitu meliputi wilayah Bogor, Bandung, Cirebon, Majalengka, Indramayu, Karawang, Bekasi, Bandung barat, Kota Bekasi. *Cluster 2* ditandai dengan warna kuning yang berarti wilayah tersebut memiliki tingkat rawan bencana alam yang rendah, yaitu meliputi wilayah Sukabumi, Cianjur, Garut, Tasikmalaya, Ciamis, Kuningan, Sumedang, Subang, Purwakarta, Pangandaran, Kota Bogor, Kota Sukabumi, Kota Bandung, Kota Cirebon, Kota Cimahi, Kota Tasikmalaya, Kota Banjar.

PENUTUP

SIMPULAN

Berdasarkan hasil penelitian analisis *cluster* wilayah rawan bencana alam menggunakan algoritma *K-Medoids* ini diperoleh bahwa algoritma *K-Medoids* dengan *Euclidean distance* memiliki 2 *cluster* dengan nilai DBI sebesar 0,305785, sedangkan *K-Medoids* dengan *Manhattan distance* memiliki 2 *cluster* dengan nilai DBI sebesar 0,299043. Sehingga berdasarkan nilai DBI, dapat diketahui bahwa *K-Medoids* dengan *Manhattan distance* merupakan hasil *clustering* yang paling optimal. Dari hasil analisis *clustering* menggunakan *K-Medoids* dengan *Manhattan distance* diperoleh 2 *cluster* yaitu, *cluster 1* merupakan wilayah dengan tingkat rawan bencana yang tinggi, dan *cluster 2* merupakan wilayah dengan tingkat rawan bencana yang rendah. Hasil analisis *cluster* ini dapat digunakan untuk melakukan perencanaan mitigasi bencana alam dan penyusunan kebijakan dalam kesiapsiagaan pada wilayah Jawa Barat.

SARAN

Pada penelitian selanjutnya, dapat diterapkan penggunaan algoritma optimasi pada algoritma *clustering* seperti *Genetic Algorithm* (GA), *Simulated Annealing* (SA), dan lainnya, untuk mengevaluasi apakah terdapat peningkatan yang signifikan dalam hasil *clustering*. Selain itu, penggunaan metrik jarak yang lain sebagai perbandingan dengan hasil penelitian ini juga dapat dipertimbangkan untuk memperoleh hasil *clustering* yang lebih baik dan optimal.

DAFTAR PUSTAKA

- Aggarwal, C. C., & K.Reddy, C. (2014). Data Clustering Algorithm and Application. In *Data Clustering*.
<https://doi.org/10.1201/9781315373515-4>
- Arora, P., Deepali, & Varshney, S. (2016). Analysis of K-Means and K-Medoids Algorithm for Big Data. *Physics Procedia*, 78(December 2015), 507–512.
<https://doi.org/10.1016/j.procs.2016.02.095>
- Everitt, B. S., & Leese, M. (2011). Cluster Analysis 5th Edition. In D. J. Balding (Ed.), *Quality and Quantity* (2011th ed., Vol. 14, Issue 1). John Wiley & Sons, Ltd.
<https://doi.org/10.1007/BF00154794>
- Grabusts, P. (2011). The choice of metrics for clustering algorithms. *Vide. Tehnologija. Resursi - Environment, Technology, Resources*, 2(1), 70–78.
<https://doi.org/10.17770/etr2011vol2.973>
- Halik, M. F. Al, & Septiana, L. (2022). Analisa Data Untuk Prediksi Daerah Rawan Bencana Alam Di Jawa Barat Menggunakan Algoritma K-Means Clustering. *Journal of Information System, Applied, Management, Accounting and Research*, 6(4), 856–870.
<https://doi.org/10.52362/jisamar.v6i4.939>
- Jefri Ardian Nugroho, Bangun Muljo Sukojo, & Inggit Lolita Sari. (2010). Pemetaan Daerah Rawan Longsor Dengan Penginderaan Jauh Dan Sistem Informasi Geografis (Studi Kasus : Kawasan Hutan Lindung Kabupaten Mojokerto). *Geoid*, 5(2), 110–117.
<https://iptek.its.ac.id/index.php/geoid/article/view/7340/4880>
- Karo, I. M., Dewi, S., Mardiana, M., Ramadhani, F., & Harliana, P. (2023). K-Means and K-Medoids Algorithm Comparison for Clustering Forest Fire Location in Indonesia. *Jurnal Ecotipe (Electronic, Control, Telecommunication, Information, and Power Engineering)*, 10(1), 86–94.
<https://doi.org/10.33019/jurnalecotipe.v10i1.3896>
- Kartikasari, M. D. (2021). Self-Organizing Map Menggunakan Davies-Bouldin Index dalam Pengelompokan Wilayah Indonesia Berdasarkan Konsumsi Pangan. *Jambura Journal of Mathematics*, 3(2), 187–196.
<https://doi.org/10.34312/jjom.v3i2.10942>
- Kaufman, L., & Rousseeuw, P. J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. A John Wiley & Sons, INC.
- Maulana Ichsan, Erlangga Alfath Wijaya, Mohammad Raffi Mahendra, & Faisal Amar Alfarouk. (2024). Pengelompokan Data Wilayah Rawan Bencana Alam di Pulau Jawa. *Neptunus: Jurnal Ilmu Komputer Dan Teknologi Informasi*, 2(3), 46–51.
<https://doi.org/10.61132/neptunus.v2i3.181>
- Meiriza, A., Ali, E., Rahmiati, & Agustin. (2023). Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Program BPJS Ketenagakerjaan. *The Indonesian Journal of Computer Science*, 12(2), 714–728.
<https://doi.org/10.33022/ijcs.v12i2.3184>
- Murdiaty, M., Angela, A., & Sylvia, C. (2020). Pengelompokan Data Bencana Alam Berdasarkan Wilayah, Waktu, Jumlah Korban dan Kerusakan Fasilitas Dengan Algoritma K-Means. *Jurnal Media Informatika Budidarma*, 4(3), 744. <https://doi.org/10.30865/mib.v4i3.2213>
- Muthmainah, S. G., Hadiana, A. I., & Melina. (2024). Comparative Analysis of K-Means and K-Medoids Clustering in Retail Store Product Grouping. 5(3), 280–294.
- Nishom, M. (2019). Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square. *Jurnal Informatika: Jurnal Pengembangan IT*, 4(1), 20–24.
<https://doi.org/10.30591/jpit.v4i1.1253>
- Sunge, A. S., Heryadi, Y., Religia, Y., & Lukas. (2020). Comparison of Distance Function to Performance of K-Medoids Algorithm for Clustering. *Proceeding - ICoSTA 2020: 2020 International Conference on Smart Technology and Applications: Empowering Industrial IoT by Implementing Green Technology for Sustainable Development*.
<https://doi.org/10.1109/ICoSTA48221.2020.1570615793>
- Sutiyo dan Eviany. (2023). Perlindungan Masyarakat. In *Angewandte Chemie International Edition*, 6(11), 951–952. (Issue Mi).