

# PERBANDINGAN ALGORITMA K-MEDOIDS DAN DBSCAN UNTUK CLUSTERING WILAYAH RAWAN BANJIR

Gregorio Melvin Karnikov<sup>1)</sup> Novario Jaya Perdana<sup>2)</sup>

<sup>1)</sup> Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas  
Fakultas Teknologi Informasi, Universitas Tarumanagara, Jln. Letjen S. Parman No.  
1, Jakarta, 11440, Indonesia

email : [gregorio.535220085@stu.untar.ac.id](mailto:gregorio.535220085@stu.untar.ac.id) <sup>1)</sup>, [novariojp@fti.untar.ac.id](mailto:novariojp@fti.untar.ac.id) <sup>2)</sup>

## ABSTRAK

Penelitian ini menganalisis tingkat kerawanan banjir di Provinsi DKI Jakarta pada tingkat kecamatan menggunakan algoritma K-Medoids dan DBSCAN. Penelitian ini dilakukan dengan menggunakan data kejadian bencana banjir yang disediakan oleh Badan Penanggulangan Bencana Daerah (BPBD) DKI Jakarta untuk melakukan pengelompokan seluruh kecamatan yang berada pada Provinsi DKI Jakarta berdasarkan kesamaan dampak banjir yang dialami. Penelitian ini mencakup pengumpulan data, pemrosesan data dan implementasi setiap algoritma. Hasil penelitian menunjukkan bahwa algoritma DBSCAN secara signifikan lebih unggul. Dengan parameter yang dioptimalkan yaitu  $\epsilon = 0.20$ ,  $MinPts = 7$ , DBSCAN mencapai Silhouette Score sebesar 0.539 dan berhasil mengidentifikasi tiga kluster kerawanan yang bermakna (Rendah, Sedang, Tinggi) serta mengisolasi data anomali (noise). Sebaliknya, konfigurasi terbaik K-Medoids ( $k=2$ ) hanya menghasilkan Silhouette Score sebesar 0.412, yang mengindikasikan struktur kluster yang lemah. Kesimpulannya, kemampuan DBSCAN dalam mendeteksi kluster dengan bentuk arbitrer dan menangani outlier menjadikannya metode yang lebih andal dan sesuai untuk analisis data kebencanaan spasial yang kompleks dan non-linier. Peta kerawanan yang dihasilkan dari klusterisasi DBSCAN dapat menjadi dasar yang kuat bagi pemangku kepentingan untuk merancang kebijakan mitigasi banjir yang lebih efektif dan tepat sasaran.

## Key words

Clustering, K-Medoids, DBSCAN, Banjir, Jakarta

## 1. Pendahuluan

Banjir merupakan peristiwa yang terjadi saat meluapnya air yang menggenangi suatu wilayah sehingga menutupi permukaan tanah yang biasanya kering.

Peristiwa ini dapat terjadi karena berbagai faktor, dimana ini bisa disebabkan akibat curah hujan yang tinggi, meluapnya sungai, kerusakan daerah aliran sungai, buruknya sistem drainase, penumpukan sampah, maupun kombinasi dari faktor-faktor tersebut. Secara alami, banjir dapat menjadi bagian dari siklus hidrologi, namun dalam konteks perkotaan dan permukiman, banjir merupakan salah satu bencana yang sering menimbulkan kerugian besar baik dari sisi ekonomi, sosial, maupun lingkungan [1].

Di provinsi DKI Jakarta, banjir merupakan salah satu bencana alam yang paling sering terjadi dan diikuti dengan kerugian yang cukup signifikan di berbagai aspek. Tingginya intensitas curah hujan, buruknya sistem drainase, banyaknya sampah berserakan dan alih fungsi lahan menjadi beberapa faktor utama yang memperparah risiko banjir di wilayah perkotaan padat seperti di kota Jakarta. Jakarta juga memiliki kondisi geografis yang didominasi daerah aliran sungai besar serta laju urbanisasi yang tinggi namun tidak diimbangi dengan tata kelola lingkungan yang baik sehingga banjir menjadi masalah rutin di kota Jakarta. Dampak dari banjir sendiri meliputi kerusakan infrastruktur, hilangnya mata pencaharian, gangguan kesehatan, hingga korban jiwa.

Di DKI Jakarta, banjir tidak hanya menimbulkan kerugian material dan gangguan aktivitas masyarakat, tetapi juga berdampak besar secara ekonomi. Kepala Badan Penanggulangan Bencana Daerah (BPBD) DKI Jakarta, Isnawa Adji, menyampaikan bahwa dampak banjir mengakibatkan kerugian mencapai Rp 2,1 triliun [2]. Nilai ini menunjukkan betapa seriusnya dampak ekonomi yang ditimbulkan oleh banjir, dan menggarisbawahi kebutuhan mendesak untuk memperkuat sistem mitigasi di wilayah perkotaan seperti Jakarta [3].

Permasalahan ini mendorong kebutuhan akan strategi mitigasi berbasis data yang dapat membantu pihak berwenang, seperti Badan Penanggulangan Bencana Daerah (BPBD) DKI Jakarta, dalam merancang kebijakan

penanggulangan banjir yang lebih efektif dan tepat sasaran. Namun, dalam praktiknya, data spasial kebencanaan sering kali bersifat tidak linier, mengandung anomali, dan memiliki distribusi yang kompleks.

## 2. Desain Penelitian

Penelitian Penelitian ini menggunakan pendekatan kuantitatif dengan menerapkan analisis kluster untuk mengelompokkan seluruh kecamatan di Provinsi DKI Jakarta berdasarkan tingkat kerawanan banjir. Pendekatan ini bertujuan untuk mengidentifikasi kelompok wilayah yang memiliki karakteristik serupa dalam hal faktor-faktor yang memengaruhi kerentanan terhadap banjir.

Melalui analisis kluster, penelitian ini diharapkan dapat mengungkap pola tersembunyi dalam data, seperti hubungan antar kecamatan yang memiliki kesamaan kondisi dampak dari banjir yang terjadi. Hasil pengelompokan ini dapat menjadi dasar bagi pemerintah dan pemangku kepentingan dalam merumuskan kebijakan mitigasi bencana yang lebih terarah dan berbasis data.

Penelitian ini akan menggunakan dua metode utama untuk melakukan pengelompokan kecamatan berdasarkan kerawanan banjir, yaitu metode K-Medoids dan DBSCAN. Kedua metode ini dipilih karena memiliki pendekatan yang berbeda secara fundamental, sehingga memungkinkan dilakukan evaluasi dari perspektif yang beragam. K-Medoids merupakan metode *clustering* berbasis partisi, sedangkan DBSCAN adalah metode *clustering* berbasis kepadatan yang mampu mengidentifikasi bentuk *cluster* yang arbitrer dan mendeteksi data anomali sebagai *noise*.

### 2.1. Algoritma K-Medoids

K-Medoids merupakan algoritma clustering yang berbasis partisi yang membagi sejumlah objek ke dalam beberapa cluster berdasarkan kesamaan karakteristik. Proses algoritma K-Medoids diawali dengan pemilihan *medoid* secara acak sebagai pusat *cluster* dan kemudian akan dilanjutkan dengan iterasi selanjutnya sampai *total cost* lebih kecil dibanding *cost* sebelumnya. *Medoid* sendiri adalah objek yang berada paling representatif atau paling tengah dalam sebuah *cluster*, sehingga metode ini lebih tahan terhadap *outlier* [4].

Tahapan untuk menerapkan algoritma K-Medoids untuk melakukan clustering adalah [5]:

1. Menentukan dan menginisialisasi jumlah kluster atau kelompok (K) yang akan membagi data.
2. Menentukan *medoid* awal secara acak sebanyak jumlah kluster yang telah ditentukan sebelumnya.
3. Menghitung jarak setiap titik data ke *medoid*, menggunakan metode seperti *Euclidean Distance* yang dapat dilihat

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \dots \dots \dots (1)$$

Di mana

1.  $d$ : Jarak antara dua titik (x dan y)
2.  $x$ : Titik / data pertama
3.  $y$ : Titik / data kedua
4.  $x_i$ : Nilai ke-i dari titik x
5.  $y_i$ : Nilai ke-I dari titik y
6.  $i$ : Indeks atribut
4. Memilih objek baru yang non *medoids* sebagai *medoid* yang baru kemudian melakukan perhitungan
5. Menghitung *total cost*, yaitu total dari jarak semua titik setiap data ke *medoid* masing-masing, sebagai pengukuran kualitas hasil pembagian kluster yang dilakukan
6. Menghitung dan mengevaluasi total simpangan (S) dari perubahan *total cost* setelah iterasi selanjutnya untuk mengganti *medoid*. Jika  $S < 0$ , pergantian *medoid* mengurangi *total cost* dan iterasi dilanjutkan. Jika  $S > 0$ , pergantian tidak menguntungkan, sehingga *medoid* tetap dan hasil klusterisasi mengikuti iterasi sebelumnya
7. Mengulangi langkah 4 sampai 6 hingga tidak terjadi lagi sampai mencapai total simpangan = 0 atau tidak ada lagi perubahan *medoid* yang menghasilkan *total cost* yang lebih rendah

### 2.2. Algoritma DBSCAN

Algoritma DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) adalah salah satu metode clustering berbasis kepadatan (*density-based*). Berbeda dengan metode seperti K-Means atau K-Medoids yang membagi data berdasarkan jarak ke pusat cluster, DBSCAN membentuk cluster berdasarkan kepadatan titik data dalam suatu area [6]

Berikut tahapan implementasi dari algoritma DBSCAN:

1. Menentukan parameter *epsilon* ( $\epsilon$ ) dan *MinPts*. Dimana parameter  $\epsilon$  merupakan radius maksimum untuk mencari tetangga dari suatu titik dan *MinPts* merupakan jumlah minimum titik dalam radius  $\epsilon$  agar sebuah titik dianggap *core point* [7].
2. Menghitung jarak setiap titik dimana tujuan dari perhitungan ini adalah untuk mengetahui seberapa dekat atau jauh setiap titik data dengan titik lainnya. Hasil perhitungan jarak inilah yang menjadi dasar dalam menentukan tetangga suatu titik. Setelah mendapatkan hasilnya maka Klasifikasi titik data berdasarkan jarak yang didapatkan. Dimana perhitungan jarak akan menggunakan *Euclidian Distance* yang dapat dilihat pada persamaan 1.
3. Mengklasifikasikan titik data setelah melalui proses perhitungan jarak dimana, titik data akan diklasifikasikan menjadi tiga kategori. Titik yang memiliki jumlah tetangga lebih dari atau sama dengan *MinPts* dalam radius  $\epsilon$  disebut

sebagai *Core Point*. Titik yang berada dalam radius  $\epsilon$  dari *core point* tetapi tidak memiliki cukup tetangga untuk menjadi *core point* dikategorikan sebagai *Border Point*. Sementara itu, titik yang tidak memenuhi kedua kriteria tersebut dianggap sebagai *Noise (Outlier)* [8].

4. Membentuk *Cluster* Proses dimana proses dimulai melalui *core point* dan diperluas dengan menambahkan titik-titik tetangga yang termasuk dalam radius  $\epsilon$ . *Cluster* terus diperbesar dengan menambahkan tetangga dari *core point* baru yang ditemukan, hingga tidak ada lagi titik yang bisa ditambahkan. Proses ini diulangi untuk semua *core point* yang belum termasuk dalam *cluster*, sehingga terbentuklah sejumlah *cluster* dengan bentuk yang bisa bervariasi, serta titik-titik yang tidak tergabung akan ditandai sebagai *noise*.

### 2.3. Algoritma DBSCAN

Metode yang akan digunakan untuk melakukan pengelompokan wilayah rawan banjir dalam penelitian ini adalah algoritma K-Medoids dan DBSCAN. Kedua metode ini dipilih karena memiliki pendekatan yang berbeda secara fundamental, sehingga memungkinkan dilakukan evaluasi dari perspektif yang beragam. K-Medoids merupakan metode *clustering* berbasis partisi, sedangkan DBSCAN adalah metode *clustering* berbasis kepadatan yang mampu mengidentifikasi bentuk cluster yang arbitrer dan mendeteksi data anomali sebagai *noise*. Perbandingan singkat dari kedua metode dapat dilihat pada Tabel 1.

Tabel 1 Perbandingan umum algoritma K-Medoids dan DBSCAN

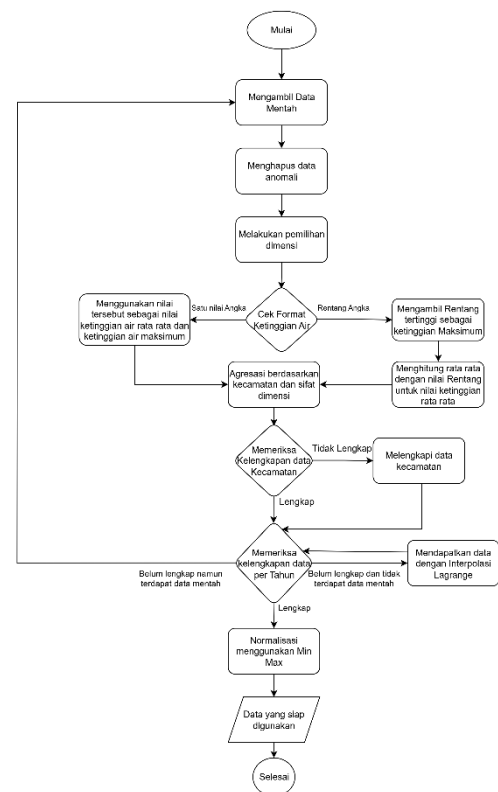
| Aspek                  | K-Medoids                        | DBSCAN   |
|------------------------|----------------------------------|--|
| Tipe <i>Clustering</i> | Partisi ( <i>Partitioning</i> )  | Berdasarkan kepadatan ( <i>Density-based</i> )                         |
| Jumlah <i>Cluster</i>  | Ditentukan di awal               | Otomatis terbentuk   |
| Parameter              | $K$ (jumlah <i>cluster</i> )     | $\epsilon$ ( $\epsilon$ ) dan $MinPts$                                 |
| Sifat                  | Sensitif terhadap <i>outlier</i> | Secara alami akan mengidentifikasi <i>outlier</i> sebagai <i>noise</i> |

### 2.4. Pengumpulan dan Pengolahan Data

Data yang digunakan dalam penelitian ini adalah data yang disediakan oleh Badan Penanggulangan Bencana Daerah (BPBD) DKI Jakarta yang diperoleh dari situs <https://satudata.jakarta.go.id/> dimana dataset yang

dikumpulkan adalah dataset Data Kejadian Bencana Banjir di DKI Jakarta yang dikumpulkan dari tahun 2019-2020 dan 2023-2025. Dimana dataset ini mencatat kejadian bencana banjir di DKI Jakarta setiap bulannya dan selalu diperbarui datanya setiap tiga bulan sekali.

Sebelum dilakukan analisis, data terlebih dahulu melalui tahap pemrosesan data (*data preprocessing*) yang mencakup beberapa langkah, antara lain pembersihan data untuk menghapus data anomali dan inkonsistensi, penyesuaian serta verifikasi data kejadian agar sesuai dengan catatan resmi per tahun dan per kecamatan, serta agregasi data untuk menggabungkan frekuensi kejadian banjir berdasarkan periode waktu dan wilayah administrasi. Tahapan ini dilakukan untuk memastikan bahwa data yang digunakan memiliki kualitas dan konsistensi yang memadai sebelum dianalisis lebih lanjut menggunakan metode klasterisasi. Dimana alur dari proses tersebut dapat dilihat pada gambar 1.



Gambar 1 Diagram alur pengumpulan dan pemrosesan data

Namun terdapat data tahun yang tak lengkap Dimana untuk tahun-tahun di mana dataset mentah tidak tersedia, akan dilakukan interpolasi Lagrange untuk memperkirakan nilai variabel yang dibutuhkan Dimana Metode ini bertujuan memperkirakan atau mendekati nilai suatu fungsi pada titik yang belum diketahui, khususnya di antara titik-titik data yang sudah tersedia (interpolasi). Konsep dasarnya adalah setiap nilai fungsi pada titik data digunakan untuk membentuk polinomial interpolasi, sehingga polinomial tersebut akan melewati seluruh titik data yang diberikan. Dengan demikian, metode ini sangat

bermanfaat dalam berbagai bidang, seperti pemrosesan data, peramalan, maupun rekayasa, karena dapat menghasilkan pendekatan yang cukup akurat meskipun hanya berdasarkan sejumlah titik data diskrit [9]. Proses interpolasi ini menggunakan dataset yang telah diproses dari tahun-tahun sebelumnya dan sesudahnya sebagai referensi, sehingga data yang dihasilkan tetap konsisten dan dapat digunakan dalam analisis *clustering*. Dimana rumus Interpolasi Lagrange dapat dilihat melalui persamaan 2.

$$P_n(x) = \sum_{i=0}^n y_i \cdot \prod_{j \neq i}^n \frac{x - x_j}{x_i - x_j} \dots\dots\dots(2)$$

Di mana

1.  $P_n(x)$  = Nilai polinomial interpolasi pada titik yang diinginkan.
2.  $y_i$  = Nilai fungsi pada titik data  $x_i$ .
3.  $x_i$  = Titik data ke-iii yang diketahui.
4.  $n$  = Derajat polinomial = jumlah titik data - 1
5.  $x$  = Titik di mana interpolasi ingin diperkirakan.

Setelah data yang telah dibutuhkan telah dikumpulkan dan didapatkan maka data siap digunakan untuk klasterisasi, Dimana variabel yang dipilih untuk digunakan dalam mengelompokkan kecamatan rawan banjir di Provinsi DKI Jakarta terdiri dari:

1. Jumlah RW Terdampak: Variabel ini menunjukkan jumlah Rukun Warga (RW) yang terdampak oleh banjir di setiap kecamatan. Variabel ini digunakan untuk mengukur skala dan sebaran wilayah yang mengalami kerawanan banjir.
2. Jumlah KK Terdampak: Variabel ini menunjukkan jumlah Kepala Keluarga (KK) yang terdampak oleh banjir di setiap kecamatan.
3. Jumlah Jiwa Terdampak: Variabel ini menunjukkan total jumlah penduduk yang terdampak oleh banjir di setiap kecamatan, dimana data ini diperoleh dari jumlah anggota keluarga (jiwa) dalam setiap Kepala Keluarga (KK) yang terdampak.
4. Ketinggian Air Rata Rata: Variabel ini menunjukkan tinggi rata-rata permukaan air saat terjadi banjir di setiap kecamatan.
5. Ketinggian Air Maksimum: Variabel ini menunjukkan tinggi maksimum permukaan air yang tercatat selama periode banjir di setiap kecamatan.

## 2.5. Metode Evaluasi

Setelah proses klasterisasi dilakukan, hasilnya akan dievaluasi dengan menggunakan beberapa metrik untuk memastikan kualitas klaster yang terbentuk. Metrik evaluasi yang digunakan dalam penelitian ini metode Sillhouette Score atau yang sering disebut juga sebagai Sillhouette Coefficient dimana metode ini akan menilai kualitas hasil *cluster* pada metode *clustering*. Metrik ini mengukur kesesuaian suatu titik data terhadap kelompoknya melalui dua aspek utama, yaitu kohesi yang menggambarkan kedekatan dan kemiripan titik data dengan anggota lain dalam *cluster* yang sama serta separasi yang menunjukkan jarak dan perbedaan titik data dengan cluster terdekat [10].

Persamaan atau rumus dari perhitungan Sillhouette Score dapat dilihat pada persamaan 3.

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \dots\dots\dots(3)$$

Di mana

1.  $S(i)$  = Silhouette Coefficient pada objek ke-i.
2.  $a(i)$  = Rata-rata jarak objek ke-i dengan anggota lain dalam cluster yang sama.
3.  $b(i)$  = Jarak rata-rata terkecil objek ke-i terhadap anggota pada cluster lain.

## 3. Hasil Percobaan

Setelah data sudah dikumpulkan dan dibersihkan maka proses klasterisasi akan dilakukan dengan membandingkan dua algoritma yaitu algoritma K-Medoids dan algoritma DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*).

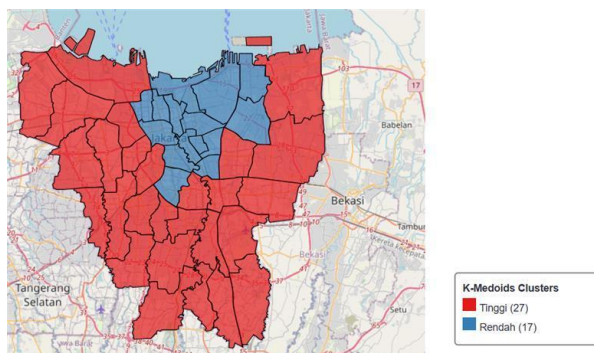
### 3.1. Algoritma K-Medoids

Algoritma K-Medoids diterapkan sebanyak sembilan kali percobaan dengan jumlah klaster yang bervariasi. Setiap percobaan dilakukan untuk mengamati perubahan hasil pengelompokan dan mengevaluasi kualitas klaster yang terbentuk pada setiap jumlah klaster. Berikut merupakan hasil percobaan klasterisasi menggunakan algoritma K-Medoids yang dapat dilihat pada Tabel 2.

Tabel 2 Laporan Evaluasi Eksperimen Algoritma K-Medoids

| Eksperimen Ke-             | Parameter (k) | Silhouette Score |
|----------------------------|---------------|------------------|
| 1                          | 2             | 0.412            |
| 2                          | 3             | 0.294            |
| 3                          | 4             | 0.264            |
| 4                          | 5             | 0.309            |
| 5                          | 6             | 0.252            |
| 6                          | 7             | 0.085            |
| 7                          | 8             | 0.195            |
| 8                          | 9             | 0.184            |
| 9                          | 10            | 0.172            |
| Rata rata Silhouette Score |               | 0.241            |

Eksperimen yang dilakukan menggunakan algoritma K-Medoids menunjukkan bahwa konfigurasi dengan jumlah kluster  $k = 2$  menghasilkan nilai Silhouette Score sebesar 0,412. Berdasarkan interpretasi umum terhadap nilai Silhouette, skor ini menunjukkan adanya struktur kluster yang lemah hingga sedang dan tidak cukup kuat untuk menyatakan pemisahan kluster yang baik (nilai Silhouette  $\geq 0,50$  biasanya dianggap menunjukkan struktur kluster yang dapat diterima, sedangkan nilai  $\geq 0,70$  menunjukkan pemisahan yang baik) [11]. Oleh karena itu, meskipun  $k = 2$  memberikan nilai Silhouette tertinggi di antara percobaan yang diuji, hasil tersebut mengindikasikan bahwa pola pengelompokan dalam data hanya memiliki kecenderungan pemisahan yang terbatas dan bahwa hasil klasterisasi harus ditafsirkan dengan hati-hati.



Gambar 2 Pemetaan Kecamatan Berdasarkan Algoritma K-Medoids

Gambar 2 menunjukkan persebaran spasial Tingkat kerawanan banjir pada kecamatan di Provinsi DKI Jakarta berdasarkan dampaknya yaitu tinggi (merah) dan rendah (biru). Peta ini dihasilkan dengan mengaitkan setiap data produksi dengan koordinat geografis provinsi, menghasilkan visualisasi spasial yang informatif.

### 3.2. Algoritma DBSCAN

Pada tahap ini, dilakukan pencarian parameter optimal untuk algoritma Density-Based Spatial Clustering of Applications with Noise (DBSCAN). Kinerja DBSCAN sangat bergantung pada dua parameter utama yaitu epsilon atau  $\epsilon$  dan jumlah minimum titik dalam radius tersebut yaitu MinPts. Penentuan kedua parameter ini krusial untuk menemukan struktur kluster yang paling alami dalam data.

Untuk mendapatkan konfigurasi terbaik, serangkaian 47 eksperimen dilakukan dengan menguji nilai Epsilon dan MinPts. Dimana hasil percobaannya dapat dilihat pada tabel 2.

Tabel 3 Laporan Evaluasi Eksperimen Algoritma DBSCAN

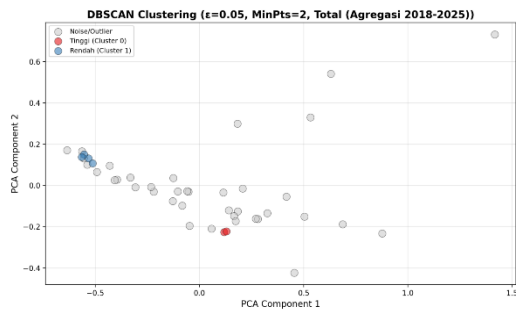
| Eksperimen Ke- | Epsilon | MinPts | Silhouette Score |
|----------------|---------|--------|------------------|
| 1              | 0.05    | 2      | 0.942            |
| 2              | 0.06    | 2      | 0.926            |

|                            |      |   |       |
|----------------------------|------|---|-------|
| 3                          | 0.07 | 2 | 0.898 |
| 4                          | 0.08 | 2 | 0.833 |
| 5                          | 0.09 | 2 | 0.701 |
| 6                          | 0.09 | 3 | 0.824 |
| 7                          | 0.10 | 2 | 0.673 |
| 8                          | 0.10 | 3 | 0.793 |
| 9                          | 0.11 | 2 | 0.625 |
| 10                         | 0.11 | 3 | 0.722 |
| 11                         | 0.12 | 2 | 0.603 |
| 12                         | 0.12 | 3 | 0.622 |
| 13                         | 0.12 | 4 | 0.714 |
| 14                         | 0.12 | 5 | 0.749 |
| 15                         | 0.13 | 2 | 0.543 |
| 16                         | 0.13 | 3 | 0.622 |
| 17                         | 0.13 | 4 | 0.714 |
| 18                         | 0.13 | 5 | 0.749 |
| 19                         | 0.14 | 2 | 0.539 |
| 20                         | 0.14 | 3 | 0.590 |
| 21                         | 0.14 | 4 | 0.658 |
| 22                         | 0.14 | 5 | 0.666 |
| 23                         | 0.15 | 2 | 0.512 |
| 24                         | 0.15 | 3 | 0.568 |
| 25                         | 0.15 | 4 | 0.617 |
| 26                         | 0.15 | 5 | 0.658 |
| 27                         | 0.16 | 2 | 0.517 |
| 28                         | 0.16 | 4 | 0.592 |
| 29                         | 0.16 | 5 | 0.633 |
| 30                         | 0.16 | 6 | 0.658 |
| 31                         | 0.17 | 2 | 0.494 |
| 32                         | 0.17 | 4 | 0.536 |
| 33                         | 0.17 | 5 | 0.556 |
| 34                         | 0.17 | 6 | 0.605 |
| 35                         | 0.18 | 2 | 0.495 |
| 36                         | 0.18 | 4 | 0.524 |
| 37                         | 0.18 | 5 | 0.535 |
| 38                         | 0.18 | 6 | 0.563 |
| 39                         | 0.18 | 7 | 0.591 |
| 40                         | 0.19 | 2 | 0.443 |
| 41                         | 0.19 | 3 | 0.495 |
| 42                         | 0.19 | 6 | 0.535 |
| 43                         | 0.19 | 7 | 0.582 |
| 44                         | 0.20 | 2 | 0.447 |
| 45                         | 0.20 | 4 | 0.458 |
| 46                         | 0.20 | 5 | 0.490 |
| 47                         | 0.20 | 7 | 0.539 |
| Rata rata Silhouette Score |      |   | 0.624 |

Dari eksperimen tersebut, dapat diidentifikasi bahwa nilai Silhouette Score tertinggi adalah 0.942, dimana nilai optimal ini dicapai pada Eksperimen Ke-1, yang menggunakan konfigurasi parameter *epsilon* sebesar 0.05 dan MinPts sebesar 2

Nilai Silhouette Score yang sangat tinggi yaitu 0.942 ini menunjukkan bahwa dengan radius yang sangat kecil dan jumlah minimum titik yang juga kecil, algoritma DBSCAN mampu menemukan kluster-kluster yang sangat padat dan terpisah dengan sangat baik namun hal

tersebut diikuti dengan banyaknya noise yang dapat dilihat pada gambar berikut.



Gambar 3 Visualisasi PDA

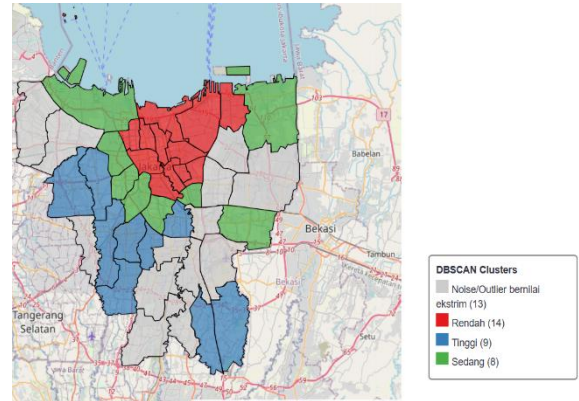
Konsekuensinya, sebagian besar data kecamatan dianggap tidak memenuhi kriteria kepadatan dan diklasifikasikan sebagai *Noise/Outlier* (titik abu-abu). Algoritma ini hanya berhasil mengelompokkan 6 titik data ke dalam dua kluster kecil: Kluster 0 (Tinggi) dan Kluster 1 (Rendah).

Perlu dipahami bahwa *Silhouette Score* pada DBSCAN umumnya dihitung hanya untuk titik-titik data yang masuk ke dalam kluster dimana tidak *noise tidak* akan terhitung.

Meskipun parameter ini berhasil menemukan dua kluster inti yang sangat padat, hasil ini tidak representatif dan tidak dapat digunakan untuk analisis komparatif yang komprehensif terhadap seluruh wilayah kecamatan. Karena tujuan penelitian adalah untuk mengelompokkan sebagian besar wilayah rawan banjir, parameter ini dianggap tidak ideal. Oleh karena itu, analisis dilanjutkan untuk mencari parameter alternatif dari tabel eksperimen yang memberikan keseimbangan lebih baik antara *Silhouette Score* yang cukup tinggi dan jumlah noise yang minimal.

Oleh karena itu, analisis diperluas untuk mencari parameter alternatif yang menyajikan keseimbangan yang lebih baik antara kualitas metrik dan inklusivitas data. Untuk memperoleh pemahaman yang lebih komprehensif mengenai struktur kluster data yang lebih menyebar, diperlukan penyesuaian parameter ke nilai *Epsilon* yang lebih besar dan *MinPts* yang lebih tinggi.

Dalam konteks ini, eksperimen Ke-47 yang menggunakan parameter *epsilon* sebesar 0.20 dan *MinPts* sebesar 7 menyajikan alternatif yang secara kualitatif lebih superior. Meskipun *Silhouette Score* menurun menjadi 0.539, konfigurasi ini memberikan hasil yang secara kontekstual jauh lebih relevan. Dengan radius yang lebih besar dan syarat kepadatan yang lebih tinggi, algoritma ini mampu mengidentifikasi 3 kluster berbeda, sekaligus secara signifikan mengurangi jumlah *noise*. Mayoritas kecamatan berhasil dipetakan ke dalam kluster yang bermakna. Dimana pemetaan tersebut dapat dilihat pada gambar 4.



Gambar 4 Pemetaan Kecamatan Berdasarkan Algoritma DBSCAN

### 3.3. Perbandingan Kinerja Algoritma K-Medoids dan Algoritma DBSCAN

Setelah proses klusterisasi dilakukan menggunakan kedua metode tersebut, hasilnya akan dievaluasi dengan menggunakan *Silhouette Score*. Untuk mendapatkan metode terbaik yang dapat digunakan. Metode terbaik akan ditentukan berdasarkan *Silhouette Score* tertinggi. Perbandingan evaluasi antara metode K-Medoids dan DBSCAN dapat dilihat melalui tabel 4.

Tabel 4 Perbandingan Evaluasi setiap algoritma

| Metode Klustering (k = 3) |                              |                  |
|---------------------------|------------------------------|------------------|
| Algoritma                 | Parameter                    | Silhouette Score |
| K-Medoids                 | k = 3                        | 0.294            |
| DBSCAN                    | Epsilon = 0.20<br>MinPts = 7 | 0.539            |

Berdasarkan tabel diatas dapat dilakukan analisis lebih lanjut terhadap karakteristik dari setiap kluster yang terbentuk. Informasi ini disajikan secara rinci pada Tabel 5 dan Tabel 6. Tabel-tabel ini menunjukkan nilai rata-rata dari setiap variabel dampak bencana banjir untuk setiap kluster, yang memberikan profil kuantitatif untuk setiap tingkat kerawanan.

Tabel 5 Karakteristik kluster DBSCAN

| Variabel                 | Rendah | Sedang  | Tinggi  | Noise    |
|--------------------------|--------|---------|---------|----------|
| Jumlah Rw Terdampak      | 33,71  | 121,38  | 209,44  | 293,69   |
| Jumlah KK Terdampak      | 60,86  | 1420,25 | 2070,11 | 8556,54  |
| Jumlah Jiwa Terdampak    | 264,36 | 5660    | 7774,22 | 29653,92 |
| Rata-rata ketinggian air | 14,03  | 26,96   | 47,59   | 46,85    |
| Ketinggian Air Maksimal  | 69,58  | 195,62  | 199,54  | 255,13   |



Tabel 6 Karakteristik kluster K-Medoids

| Variabel                 | Rendah | Sedang  | Tinggi   |
|--------------------------|--------|---------|----------|
| Jumlah Rw Terdampak      | 33,71  | 118,18  | 282,84   |
| Jumlah KK Terdampak      | 60,86  | 2361,27 | 6066     |
| Jumlah Jiwa Terdampak    | 264,36 | 8088,36 | 21672,47 |
| Rata-rata ketinggian air | 14,03  | 26,96   | 48,94    |
| Ketinggian Air Maksimal  | 69,58  | 183,56  | 245,18   |

Dari Tabel 5, terlihat bahwa kluster "Tinggi" yang diidentifikasi oleh DBSCAN memiliki karakteristik dampak yang sangat parah, baik dari segi skala (rata-rata 209 RW terdampak) maupun intensitas (ketinggian air maksimal rata-rata 199.54 cm). Menariknya, kecamatan yang diklasifikasikan sebagai "Noise" memiliki nilai dampak yang lebih ekstrem lagi yang mengonfirmasi bahwa DBSCAN berhasil mengisolasi kasus-kasus paling parah sebagai anomali yang memerlukan perhatian khusus. Sebaliknya, kluster "Tinggi" dari K-Medoids yang dapat dilihat pada Tabel 5 mencampurkan beberapa karakteristik yang oleh DBSCAN dipisahkan antara kluster "Tinggi" dan "Noise", yang menunjukkan kemampuan diferensiasi yang lebih rendah.

#### 4. Kesimpulan

Berdasarkan hasil evaluasi terhadap algoritma K-Medoids dan DBSCAN dapat disimpulkan bahwa algoritma DBSCAN secara signifikan mengungguli K-Medoids. DBSCAN tidak hanya menghasilkan skor evaluasi yang lebih tinggi tetapi juga mampu mengidentifikasi kluster dengan bentuk spasial yang kompleks dan secara efektif mengisolasi kecamatan-kecamatan dengan tingkat dampak paling ekstrem sebagai noise, yang memberikan nilai tambah signifikan untuk analisis risiko. Kontribusi utama dari penelitian ini adalah memberikan bukti empiris yang kuat bahwa untuk analisis data kebencanaan spasial yang kompleks dan non-linier, algoritma berbasis kepadatan seperti DBSCAN secara metodologis lebih sesuai dan menghasilkan wawasan yang lebih andal serta dapat ditindaklanjuti dibandingkan dengan algoritma berbasis partisi seperti K-Medoids namun, efektivitas DBSCAN bergantung pada pemilihan parameter ( $\epsilon$ ) dan  $MinPts$ ) sehingga hal tersebut perlu diperhatikan untuk mendapatkan hasil pengelompokan yang optimal. Hasil penelitian ini diharapkan dapat memberikan kontribusi nyata bagi Pemerintah Provinsi DKI Jakarta serta instansi terkait seperti Badan Penanggulangan Bencana Daerah (BPBD) dalam upaya pencegahan, penanggulangan, dan perumusan strategi penanganan banjir di wilayah DKI Jakarta. Melalui informasi mengenai pola dan karakteristik tingkat kerawanan banjir yang dihasilkan,

diharapkan hasil ini dapat menjadi bahan pertimbangan dalam pengambilan keputusan dan perencanaan kebijakan mitigasi bencana yang lebih efektif dan berbasis data.

Selain itu, hasil penelitian ini juga diharapkan bermanfaat bagi masyarakat umum, lembaga pendidikan, dan komunitas sosial dalam meningkatkan kesadaran serta kesiapsiagaan terhadap risiko banjir di lingkungan masing-masing. Dengan demikian, penelitian ini tidak hanya mendukung penyusunan kebijakan pemerintah, tetapi juga mendorong partisipasi masyarakat dalam membangun sistem mitigasi banjir yang berkelanjutan, kolaboratif, dan adaptif terhadap perubahan lingkungan di masa mendatang.

#### REFERENSI

- [1] Geosains.id, "Sungai Meluap Jadi Pemicu Banjir! Ketahui Penyebab dan Mitigasinya," 13 03 2025. [Online]. Available: <https://geosains.id/sungai-meluap-jadi-pemicu-banjir-ketahui-penyebab-dan-mitigasinya/Referensi> memiliki kebaruan maksimal 5 tahun lamanya dari tanggal penelitian
- [2] [2]. Isnawa Adji (BPBD DKI Jakarta), "BPBD Ungkap Dampak Banjir Jakarta Timbulkan Kerugian hingga Rp 2,1 Triliun," 29 05 2024. [Online]. Available: <https://news.detik.com/berita/d-7363617/bpbd>
- [3] A. Taryana, M. R. E. Mahmudi, and H. Bekt, "ANALISIS KESIAPSIAGAAN BENCANA BANJIR DI JAKARTA," JANE - Jurnal Administrasi Negara, vol. 13, no. 2, p. 302, Feb. 2022, doi: 10.24198/jane.v13i2.37997.
- [4] A. Triansyah, D. E. Herwindiati, and J. Hendryli, "PERBANDINGAN K-MEANS DAN K-MEDOIDS UNTUK KLASSTERING TINGKAT STRES PADA MANUSIA," Jurnal Ilmu Komputer Dan Sistem Informasi, vol. 10, no. 2, Aug. 2022, doi: 10.24912/jiksi.v10i2.22534.
- [5] H. D. Tampubolon, S. Suhada, M. Safii, S. Solikhun, and D. Suhendro, "Penerapan Algoritma K-Means dan K-Medoids Clustering untuk Mengelompokkan Tindak Kriminalitas Berdasarkan Provinsi", IKOMTI, vol. 2, no. 2, pp. 6–12, Nov. 2021.
- [6] I. N. Simbolon and P. D. Friskila, "ANALISIS DAN EVALUASI ALGORITMA DBSCAN (DENSITY-BASED SPATIAL CLUSTERING OF APPLICATIONS WITH NOISE) PADA TUBERKULOSIS", JITET, vol. 12, no. 3S1, Oct. 2024.
- [7] D. Deng, "DBSCAN Clustering Algorithm Based on Density," 2020 7th International Forum on Electrical Engineering and Automation (IFEEA), Hefei, China, 2020, pp. 949–953, doi: 10.1109/IFEEA51475.2020.00199.
- [8] Y. Hasan, "Pengukuran Silhouette Score dan Davies-Bouldin Index pada Hasil Cluster K-Means dan Dbscan," Jurnal Informatika dan Teknik Elektro Terapan, vol. 12, 2024.
- [9] N. N. Mansyur, N. Arman, N. La Gubu, W. Somayasa, and N. Aswani, "PENERAPAN METODE INTERPOLASI LAGRANGE DALAM MERAMALKAN JUMLAH PENDAPATAN PADA PERCETAKAN (STUDI KASUS: GEVIRA ADVERTISING)," Jurnal Matematika Komputasi Dan Statistika, vol. 4, no. 1, pp. 540–546, Jul. 2024, doi: 10.33772/jmks.v4i1.80
- [10] T. Rahmawati, Y. Wilandari, and P. Kartikasari, "ANALISIS PERBANDINGAN SILHOUETTE

COEFFICIENT DAN METODE ELBOW PADA PENGELOMPOKKAN PROVINSI DI INDONESIA BERDASARKAN INDIKATOR IPM DENGAN K-MEDOIDS,” Jurnal Gaussian, vol. 13, no. 1, pp. 13–24, Aug. 2024, doi: 10.14710/j.gauss.13.1.13-24.

- [11] M. Shutaywi and N. N. Kachouie, “Silhouette Analysis for Performance Evaluation in Machine Learning with Applications to Clustering,” Entropy, vol. 23, no. 6, p. 759, Jun. 2021, doi: 10.3390/e23060759.

**Gregorio Melvin Karnikov**, Mahasiswa S1 Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara, Jakarta.