

Pengembangan Algoritma Hybrid K-Means DBSCAN untuk Optimasi Cluster Daerah Rawan Longsor berbasis Web-Gis

Development of Hybrid K-Means DBSCAN Algorithm for Optimization of Landslide-Prone Area Clusters based on Web-GIS

¹Dede Irmayanti*, ²Teguh Iman Hermanto

^{1,2}Program Studi Teknik Informatika, Sekolah Tinggi Teknologi Wastukancana

^{1,2}Jl. Cikopak No.53, Kabupaten Purwakarta, Jawa Barat, Indonesia

*e-mail: dedeirmayanti@wastukancana.ac.id

(received: 6 August 2025, revised: 11 November 2025, accepted: 12 November 2025)

Abstrak

Bencana tanah longsor merupakan salah satu ancaman geologis utama di Provinsi Jawa Barat yang berdampak serius terhadap kehidupan sosial, ekonomi, dan infrastruktur masyarakat. Permasalahan utama dalam mitigasi longsor terletak pada ketidakakuratan dalam mengklasifikasikan wilayah rawan secara spasial dan temporal, serta keterbatasan pendekatan tunggal dalam analisis data bencana. Penelitian ini bertujuan untuk mengembangkan model klasifikasi wilayah rawan longsor berbasis data dengan pendekatan *hybrid clustering* yang menggabungkan algoritma *K-Means* dan *DBSCAN*. Data yang digunakan terdiri dari catatan kejadian longsor dari tahun 2020 hingga 2024 serta data spasial administratif kabupaten/kota. Tahapan analisis mencakup integrasi dan normalisasi data, eksplorasi statistik, penerapan *clustering K-Means* sebagai kerangka segmentasi global, serta *DBSCAN* untuk identifikasi pola lokal dan *outlier*. Validasi model dilakukan menggunakan metrik internal yaitu *Silhouette Coefficient* sebesar 0.448 dan *Davies-Bouldin Indeks* sebesar 0.602, yang menunjukkan bahwa metode *hybrid* memberikan performa terbaik dalam hal kepadatan dan pemisahan *cluster*. Hasil klasifikasi kemudian divisualisasikan dalam platform *Web-GIS* interaktif berbasis *Streamlit* dan *Folium*, yang memungkinkan pengguna memilih tahun dan metode klasifikasi serta menampilkan strategi mitigasi berdasarkan kategori risiko. Penelitian ini menyimpulkan bahwa pendekatan *hybrid clustering* mampu meningkatkan akurasi klasifikasi wilayah rawan bencana dan memberikan kontribusi signifikan dalam penyediaan informasi spasial yang lebih adaptif dan aplikatif bagi pengambilan kebijakan mitigasi di daerah rawan longsor.

Kata kunci: validasi cluster, DBSCAN, hybrid clustering, k-means, web-GIS,

Abstract

Landslides represent one of the major geological hazards in West Java Province, posing serious impacts on social life, economic activities, and public infrastructure. A key challenge in landslide mitigation lies in the inaccuracy of spatial and temporal classification of landslide-prone areas, as well as the limitations of single-method approaches in disaster data analysis. This study aims to develop a data-driven classification model for landslide-prone areas using a hybrid clustering approach that combines the K-Means and DBSCAN algorithms. The dataset consists of landslide incident records from 2020 to 2024 and administrative spatial data at the regency/city level. The analysis stages include data integration and normalization, statistical exploration, the application of K-Means clustering as a global segmentation framework, and DBSCAN for identifying local patterns and outliers. Model validation was conducted using internal evaluation metrics, yielding a Silhouette Coefficient of 0.448 and a Davies–Bouldin Index of 0.602, indicating that the hybrid method provides superior performance in terms of cluster compactness and separation. The classification results are visualized through an interactive Web-GIS platform developed using Streamlit and Folium, enabling users to select specific years and classification methods while displaying mitigation strategies based on risk categories. This study concludes that the hybrid clustering approach enhances the accuracy of landslide-prone area classification and makes a significant contribution to the provision of more adaptive and practical spatial information to support mitigation policy decision-making in landslide-vulnerable regions.

Keywords: cluster validation, DBSCAN, hybrid clustering, k-means, web-GIS

1 Pendahuluan

Indonesia merupakan negara yang rawan terhadap berbagai jenis bencana alam karena kondisi geologis, geografis, dan klimatologisnya yang kompleks. Salah satu bencana yang sering terjadi adalah tanah longsor, terutama di wilayah dengan topografi curam, struktur tanah yang labil, dan curah hujan tinggi. Provinsi Jawa Barat merupakan salah satu provinsi dengan tingkat kerawanan longsor tertinggi di Indonesia [1]. Berdasarkan data Badan Nasional Penanggulangan Bencana (BNPB), dalam lima tahun terakhir, Jawa Barat secara konsisten menempati posisi atas dalam jumlah kejadian longsor, yang mengakibatkan korban jiwa, kerusakan infrastruktur, dan gangguan terhadap aktivitas sosial ekonomi masyarakat.

Kondisi ini menunjukkan perlunya sistem deteksi dan klasifikasi wilayah rawan longsor yang lebih efektif dan responsif. Salah satu pendekatan yang berkembang dalam bidang *data science* adalah metode *clustering*, yang digunakan untuk mengelompokkan data ke dalam kategori-kategori berdasarkan kesamaan karakteristik. Dalam konteks kebencanaan, *clustering* dapat digunakan untuk mengelompokkan wilayah berdasarkan kerentanan atau tingkat risiko bencana [2]. Metode yang umum digunakan seperti *K-Means* memiliki keunggulan dalam kecepatan dan efisiensi pengelompokan berbasis *centroid*, namun lemah dalam mengidentifikasi outlier atau data dengan nilai ekstrem. Sementara itu, *DBSCAN* (*Density-Based Spatial Clustering of Applications with Noise*) unggul dalam mendeteksi *outlier* dan struktur *cluster* berdensitas tinggi, tetapi sensitif terhadap penentuan parameter dan dapat menghasilkan cluster yang tidak stabil jika diterapkan pada data berdimensi tinggi tanpa *preprocessing* [3].

Permasalahan yang ingin dipecahkan dalam penelitian ini adalah bagaimana mengembangkan pendekatan pengelompokan data yang lebih adaptif dan akurat dalam mengidentifikasi wilayah rawan longsor, dengan tetap mempertimbangkan adanya variasi ekstrem dan distribusi data yang tidak homogen. Untuk itu, penelitian ini mengusulkan pengembangan algoritma *hybrid clustering* yang menggabungkan keunggulan dari *K-Means* dan *DBSCAN*. Dalam pendekatan ini, *K-Means* digunakan untuk membentuk *cluster* makro berdasarkan distribusi global data, sementara *DBSCAN* digunakan untuk mendeteksi *outlier* dan memperhalus klasifikasi lokal berdasarkan kepadatan data[4]. Hasil kombinasi ini menghasilkan klasifikasi wilayah yang lebih realistik dan mempertimbangkan baik struktur global maupun anomali lokal dari data bencana.

Secara *state of the art*, sejumlah penelitian sebelumnya telah menerapkan *K-Means* maupun *DBSCAN* secara terpisah dalam analisis spasial kebencanaan, namun jarang dilakukan penggabungan kedua algoritma dalam satu pendekatan terpadu. Selain itu, belum banyak penelitian yang mengintegrasikan hasil analisis clustering ke dalam sistem *Web-GIS* interaktif yang dapat diakses oleh pemangku kepentingan secara *real-time* [5]. Kebaruan (*novelty*) dalam penelitian ini terletak pada pendekatan *hybrid* yang menggabungkan *K-Means* dan *DBSCAN* secara komplementer untuk optimasi klasifikasi wilayah rawan longsor, dan pengembangan platform *Web-GIS* berbasis *Streamlit-Folium* yang memungkinkan visualisasi spasial hasil *clustering* secara interaktif dan terfilter per tahun serta metode.

Tujuan dari penelitian ini adalah untuk mengembangkan sistem klasifikasi wilayah rawan longsor berbasis *hybrid clustering* dan menerapkannya dalam bentuk aplikasi *Web-GIS* yang interaktif. Manfaat dari penelitian ini diharapkan mampu mendukung pengambilan keputusan pemerintah daerah dan lembaga kebencanaan dalam melakukan mitigasi berbasis data dan spasial, sekaligus meningkatkan akses publik terhadap informasi kebencanaan secara visual dan dinamis.

2 Tinjauan Literatur

Simon, Zheng, dkk melakukan penelitian dampak bencana alam pada profil beban listrik menggunakan algoritma pengelompokan *k-means*, dengan studi kasus sistem kelistrikan Lombok di Indonesia. Studi ini menggunakan algoritma pengelompokan *k-means* untuk mengidentifikasi profil beban listrik yang terdampak bencana alam. Studi ini menemukan bahwa algoritma *k-means clustering* dapat mengidentifikasi profil beban listrik yang terdampak bencana alam, dan dampak jangka pendek dari gempa bumi dan gangguan listrik pada sistem kelistrikan tidak jauh berbeda, tetapi dampak jangka panjangnya memiliki perbedaan yang signifikan [6].

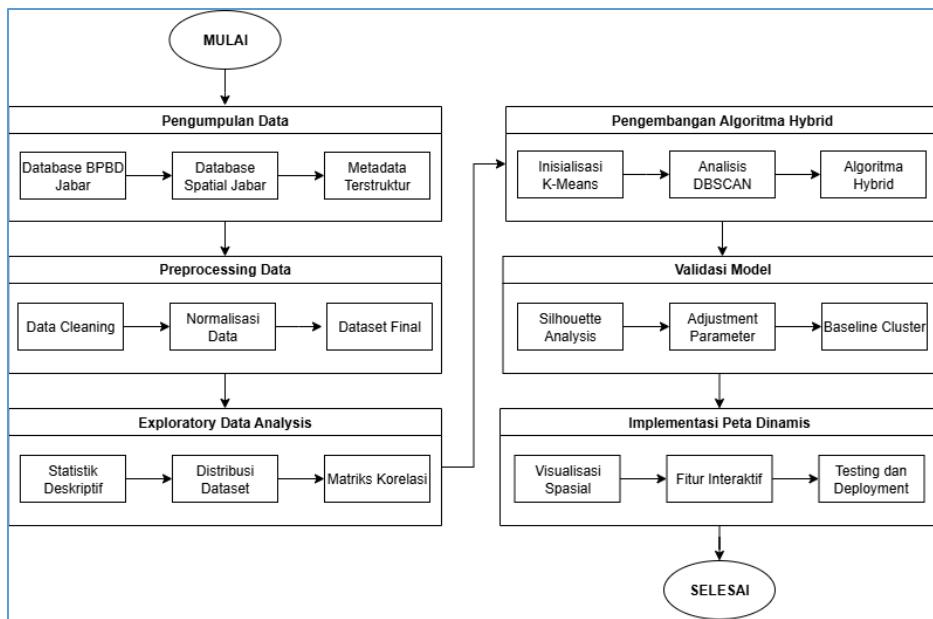
Sementara itu, studi oleh Reza, Changkye, dkk melakukan penelitian optimasi lokasi geospasial tempat penampungan darurat untuk kondisi bencana banjir. Studi ini menemukan bahwa sebagian besar penelitian yang ada menerapkan jarak langsung (Jarak *Euclidean*) dalam proses analisis, tetapi jarak tempuh dapat lebih mewakili proses evakuasi yang sebenarnya. Studi ini mengusulkan penerapan metode pengelompokan *K-Means* sebagai solusinya. Metode yang digunakan dalam penelitian ini meliputi analisis geospasial, metodologi optimasi seperti pemrograman linier dan metode *clustering*, serta penggunaan variabel keputusan seperti jarak langsung dan jarak tempuh. Penelitian ini menyimpulkan bahwa metode yang diusulkan secara umum dapat menggambarkan proses evakuasi dan pengelompokan *K-Means* dapat digunakan untuk mengoptimalkan distribusi korban bencana ke tempat penampungan darurat [7].

Nurfidah, Siti, dkk melakukan penelitian dengan membandingkan kinerja lima algoritma *clustering* (*KMedoids*, *K-Means*, *DBSCAN*, *Fuzzy C-Means*, dan *K-Affinity Propagation*) dalam mengelompokkan data gempa bumi di Indonesia dari Januari 2017 hingga Januari 2023. Hasil penelitian menunjukkan bahwa *K-Means* memberikan keseimbangan terbaik, membentuk enam klaster dengan Skor *Silhouette* sebesar 0,3245 dan Kemurnian Klaster sebesar 0,7366, sedangkan *K-Medoids* mengikutinya dengan Skor *Silhouette* sebesar 0,3158 dan Kemurnian Klaster sebesar 0,7190. Studi ini menemukan bahwa algoritma *K-Means* memiliki kinerja terbaik dalam pengelompokan data gempa bumi [8].

Berdasarkan tinjauan terhadap literatur tersebut, dapat disimpulkan bahwa belum ada penelitian yang secara komprehensif menggabungkan keunggulan *K-Means* dan *DBSCAN* dalam satu pendekatan *hybrid* untuk pengelompokan wilayah rawan longsor yang bersifat adaptif dan presisi tinggi. Selain itu, integrasi hasil *clustering* ke dalam *Web-GIS* interaktif dengan fitur filter tahun, metode, dan informasi mitigasi juga belum pernah dilakukan secara spesifik untuk wilayah Jawa Barat. Dengan demikian, penelitian ini hadir sebagai jawaban atas kekosongan tersebut dengan menawarkan pendekatan *hybrid clustering* berbasis *K-Means* dan *DBSCAN* yang langsung terintegrasi dalam aplikasi *Web-GIS* yang ringan dan mudah digunakan. Kebaruan utama terletak pada kombinasi metode untuk optimasi hasil klasifikasi spasial, deteksi risiko ekstrem berbasis *outlier*, dan pengembangan sistem visualisasi yang dapat diakses secara real-time oleh para pemangku kepentingan.

3 Metode Penelitian

Alur sistematis dari tahapan penelitian ini mulai dari proses pengumpulan data hingga pengembangan sistem *Web-GIS* interaktif. Rangkaian proses ini disusun untuk memastikan bahwa setiap langkah memiliki dasar metodologis yang kuat serta keterkaitan fungsional antara komponen data, analisis, dan aplikasi. Alur penelitian terbagi menjadi enam tahapan utama yang dimulai dari integrasi data kebencanaan dan spasial sebagai fondasi, dilanjutkan dengan tahapan *preprocessing* untuk memastikan kualitas dan keterpaduan data, serta eksplorasi awal guna mengenali karakteristik dan pola data. Selanjutnya, algoritma *hybrid clustering* dikembangkan dengan pendekatan kombinasi *K-Means* dan *DBSCAN* untuk menghasilkan klasifikasi wilayah yang lebih adaptif terhadap distribusi data yang kompleks. Validasi model dilakukan untuk menjamin konsistensi hasil *clustering*, yang kemudian ditransformasikan ke dalam visualisasi spasial dinamis melalui implementasi *Web-GIS*. Setiap tahapan dirancang untuk saling menunjang secara iteratif guna mencapai tujuan utama, yaitu menghasilkan sistem klasifikasi risiko longsor yang akurat, kontekstual, dan dapat digunakan oleh pemangku kepentingan secara langsung. Alur proses penelitian ini disajikan pada Gambar 1.



Gambar 1 Metode penelitian

- 1) Tahapan pertama adalah pengumpulan data, yang mencakup integrasi data kejadian bencana dari BPBD Provinsi Jawa Barat, data spasial wilayah administratif dalam format *GeoJSON*, serta metadata terstruktur dari sumber-sumber pendukung lainnya. Seluruh data kemudian dikompilasi menjadi basis informasi yang siap untuk diproses secara numerik dan spasial [9].
- 2) Tahap selanjutnya adalah *preprocessing data*, yang terdiri dari proses data cleaning untuk menghilangkan duplikasi dan menangani nilai hilang, serta normalisasi nilai untuk menyeragamkan skala antar variabel [10]. Data yang telah dibersihkan dan dinormalisasi ini kemudian membentuk dataset final yang siap dianalisis.
- 3) Pada tahap *exploratory data analysis (EDA)*, dilakukan analisis statistik deskriptif untuk memahami distribusi *dataset*, mengidentifikasi potensi anomali, dan membangun matriks korelasi antar variabel [11]. Analisis ini bertujuan untuk mengenali struktur awal dari data serta memilih variabel-variabel yang relevan untuk dimasukkan ke dalam model *clustering*.
- 4) Tahap inti dari penelitian ini adalah pengembangan algoritma hybrid, yang dimulai dengan inisialisasi model *K-Means* sebagai *baseline clustering*, diikuti dengan penerapan algoritma *DBSCAN* untuk mendeteksi outlier atau pola distribusi tidak homogen. Hasil dari kedua algoritma tersebut kemudian digabungkan untuk membentuk label *hybrid cluster*, yang mempertimbangkan baik pola umum maupun ekstrem [12].
- 5) Setelah proses *clustering*, dilakukan validasi model melalui analisis *Silhouette Score* untuk menilai kohesi dan separasi antar *cluster*. Jika hasil validasi belum optimal, dilakukan penyesuaian parameter (*adjustment*) seperti jumlah cluster (K), *epsilon*, dan *min_samples*, hingga diperoleh struktur *cluster* terbaik yang menjadi *baseline cluster* [13].
- 6) Tahap akhir adalah implementasi peta dinamis dalam bentuk aplikasi *Web-GIS* interaktif. Hasil *clustering* diintegrasikan ke dalam visualisasi spasial dengan fitur interaktif seperti *filter* per tahun dan *metode clustering*. Sistem ini kemudian diuji dan disiapkan untuk *deployment* agar dapat diakses secara publik atau oleh instansi terkait.
- 7) Rangkaian tahapan dalam diagram tersebut menunjukkan bahwa pendekatan yang dilakukan tidak hanya fokus pada analisis algoritmik, tetapi juga menekankan pada validasi dan keterpakaian praktis dalam bentuk sistem visual yang informatif. Hal ini mendukung tujuan utama penelitian, yakni memberikan solusi berbasis data dan spasial untuk mitigasi bencana longsor di wilayah Jawa Barat.

4 Hasil dan Pembahasan

Hasil penelitian ini disajikan secara terstruktur sesuai dengan enam tahapan metodologi yang telah dijelaskan sebelumnya: mulai dari pengumpulan data hingga implementasi *Web-GIS*. Setiap

tahap menghasilkan luaran spesifik yang mendukung pembangunan sistem klasifikasi wilayah rawan longsor yang adaptif, valid, dan aplikatif.

4.1 Pengumpulan dan Integrasi Data

Tahapan awal dalam penelitian ini adalah pengumpulan dan integrasi data yang terdiri dari dua jenis sumber utama yaitu data numerik kejadian bencana longsor dan data spasial wilayah administratif. Data numerik diperoleh dari BPBD Provinsi Jawa Barat dalam bentuk file tabular. *Dataset* ini mencakup informasi kejadian longsor yang terjadi di 27 kabupaten/kota selama periode 2020 hingga 2024. Setiap entri berisi variabel penting yang menggambarkan tingkat kerawanan bencana, antara lain jumlah kejadian longsor, jumlah korban meninggal, korban tertimbun, jumlah rumah rusak, dan total populasi terdampak. Setiap baris juga dilengkapi atribut tahun dan nama kabupaten/kota, yang menjadi kunci utama untuk integrasi spasial. Sample dataset disajikan pada Tabel 1.

Tabel 1 Data bencana longsor jawa barat

Kabupaten	Tahun	Jumlah longsor	Jiwa terdampak	Jiwa meninggal	Rusak terdampak	Rusak ringan	Rusak sedang	Rusak berat	Tertimbun
Kabupaten bogor	2020	70	890	15	169	80	56	33	44
Kabupaten sukabumi	2020	74	221	0	71	17	25	29	8
Kabupaten cianjur	2020	16	472	0	71	40	4	27	108
Kabupaten bandung	2020	16	65	0	14	3	9	2	1
Kabupaten garut	2020	51	1790	1	240	79	50	111	47

Sumber kedua adalah file *GeoJSON* wilayah administratif yang memuat batas-batas *polygon* dari seluruh kabupaten/kota di Jawa Barat yang disajikan pada Gambar 2. File ini menyediakan komponen spasial berupa koordinat geometri dan atribut identitas wilayah, seperti nama daerah dan kode wilayah. Data ini bersifat statis dan digunakan sebagai referensi visual dalam pemetaan hasil analisis clustering.



Gambar 2 wilayah administratif jawa barat

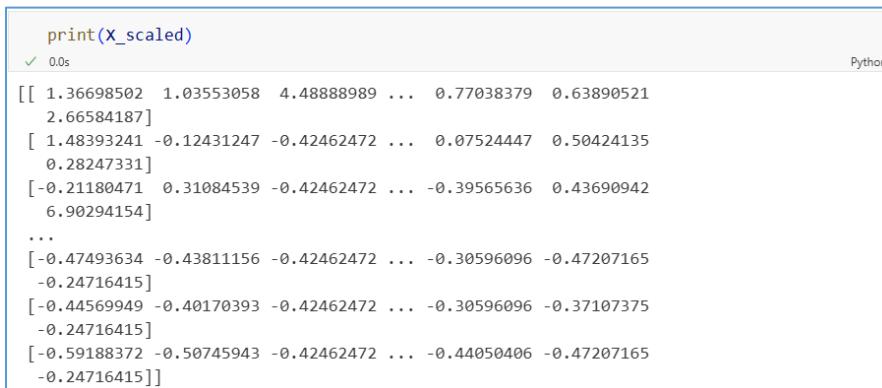
4.2 *Preprocessing* dan Transformasi Data

Langkah pertama adalah pembersihan data (*data cleaning*). Proses ini meliputi pemeriksaan nilai hilang (*missing values*), duplikasi entri, serta kesalahan input seperti nilai negatif pada jumlah korban atau rumah rusak. Ditemukan beberapa entri dengan nilai nol pada seluruh indikator, yang kemudian dianalisis sebagai data tidak representatif dan dikeluarkan dari proses analisis lanjut [14]. Selanjutnya dilakukan proses seleksi fitur untuk menentukan fitur apa saja yang akan diproses menggunakan algoritma *clustering*. Hasil data preprocessing disajikan pada Tabel 2.

Tabel 2 Hasil *preprocessing* data

Jumlah longsor	Jiwa terdampak	Jiwa meninggal	Rusak terdampak	Rusak ringan	Rusak sedang	Rusak berat	Tertimbun
70	890	15	169	80	56	33	44
74	221	0	71	17	25	29	8
16	472	0	71	40	4	27	108
16	65	0	14	3	9	2	1
51	1790	1	240	79	50	111	47

Langkah kedua adalah normalisasi nilai, dilakukan menggunakan teknik *Standard Scaler*. Proses ini mentransformasi setiap fitur sehingga memiliki distribusi dengan rata-rata nol ($\mu = 0$) dan simpangan baku satu ($\sigma = 1$). Teknik ini dipilih karena skala antar variabel dalam *dataset* sangat bervariasi, misalnya jumlah korban bisa mencapai ribuan sementara jumlah kejadian bisa di bawah 10. Penggunaan *Standard Scaler* memungkinkan algoritma *clustering* berbasis jarak seperti *K-Means* dan *DBSCAN* untuk memperlakukan semua variabel secara setara tanpa bias terhadap variabel berskala besar [15]. Hasil standarisasi data disajikan pada gambar 3.



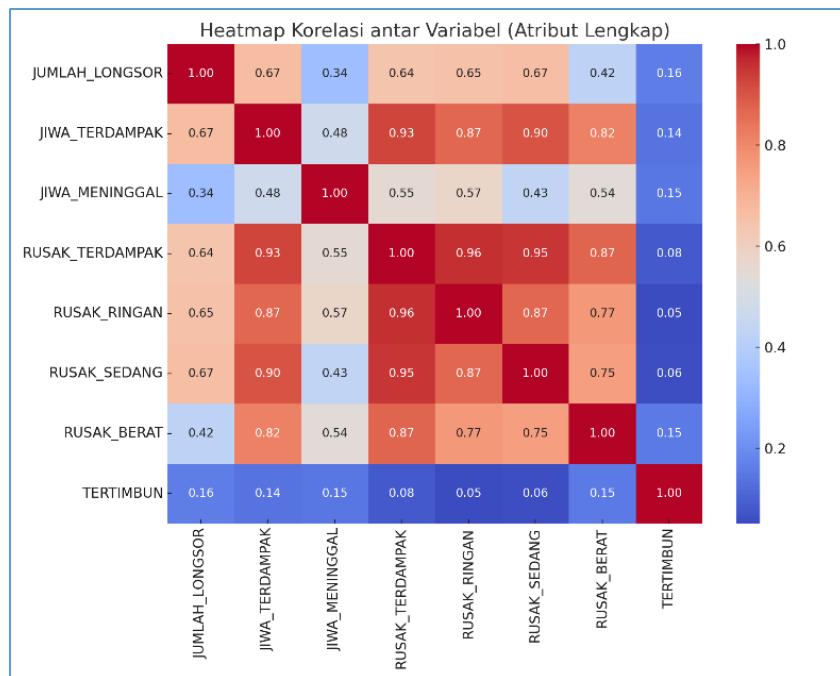
```
print(X_scaled)
0.0s
[[ 1.36698502  1.03553058  4.48888989 ...  0.77038379  0.63890521
  2.66584187]
 [ 1.48393241 -0.12431247 -0.42462472 ...  0.07524447  0.50424135
  0.28247331]
 [-0.21180471  0.31084539 -0.42462472 ... -0.39565636  0.43690942
  6.90294154]
 ...
 [-0.47493634 -0.43811156 -0.42462472 ... -0.30596096 -0.47207165
 -0.24716415]
 [-0.44569949 -0.40170393 -0.42462472 ... -0.30596096 -0.37107375
 -0.24716415]
 [-0.59188372 -0.50745943 -0.42462472 ... -0.44050406 -0.47207165
 -0.24716415]]
```

Gambar 3 Hasil standarisasi data

4.3 Exploratory Data Analysis (EDA)

Tahap eksplorasi data bertujuan untuk memahami karakteristik awal dari *dataset*, mengenali distribusi variabel, mendekripsi potensi *outlier*, serta mengidentifikasi hubungan antar fitur yang dapat memengaruhi hasil *clustering*. Proses ini penting dilakukan sebelum penerapan algoritma *machine learning*, agar arah segmentasi wilayah memiliki dasar deskriptif dan statistik yang kuat [16].

Untuk mendapatkan pemahaman lebih mendalam terhadap hubungan antar variabel, dilakukan analisis korelasi terhadap delapan atribut utama yang relevan dengan dampak bencana. Hasil visualisasi pada Gambar 4 menunjukkan bahwa JUMLAH_LONGSOR memiliki korelasi positif yang kuat terhadap RUSAK_TERDAMPAK ($r = 0.79$) dan JIWA_TERDAMPAK ($r = 0.72$). Selain itu, terdapat korelasi tinggi antara RUSAK_RINGAN, RUSAK_SEDANG, dan RUSAK_BERAT, yang menggambarkan kesesuaian distribusi skala kerusakan fisik terhadap total dampak bencana. Korelasi sedang terlihat antara JIWA_MENINGGAL dan TERTIMBUN, yang secara logis menunjukkan hubungan antara korban jiwa dan tingkat tertimbunnya penduduk. Variabel RUSAK_TERDAMPAK juga memiliki korelasi lintas kategori yang kuat terhadap semua tipe kerusakan. Secara umum, pola korelasi ini memperkuat pemilihan fitur untuk proses *clustering*, sekaligus menunjukkan bahwa variabel dampak fisik dan sosial bencana saling terhubung erat [17].

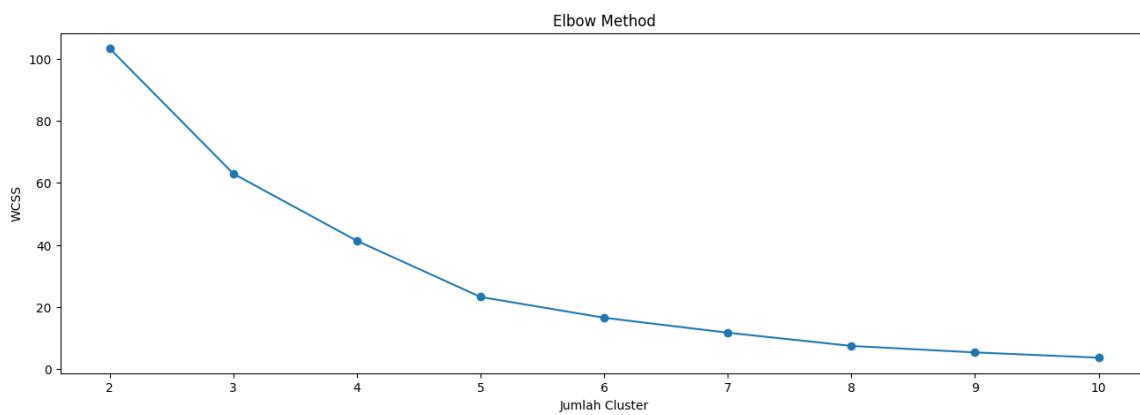


Gambar 4 Matriks korelasi dataset

4.4 Pengembangan Algoritma *Hybrid Clustering*

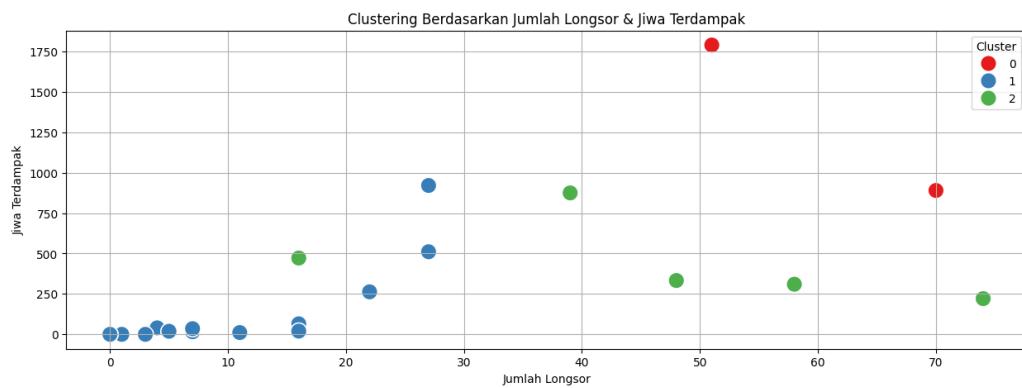
Tahapan inti dalam penelitian ini adalah pengembangan metode *Hybrid Clustering* dengan menggabungkan dua pendekatan berbeda: *K-Means* sebagai teknik partisi berbasis centroid dan *DBSCAN* sebagai teknik density-based yang adaptif terhadap distribusi data tidak beraturan. Tujuannya adalah menghasilkan klasifikasi wilayah rawan bencana longsor yang tidak hanya stabil secara global, tetapi juga sensitif terhadap pola lokal dan *outlier*.

Proses diawali dengan penerapan algoritma *K-Means* pada data yang telah dinormalisasi. Metode *Elbow* digunakan untuk menentukan nilai K yang optimal dengan memetakan nilai *Within Cluster Sum of Squares* (WCSS) terhadap jumlah cluster. Hasil analisis pada Gambar 5 menunjukkan bahwa titik siku (*elbow*) berada pada K = 3, yang berarti pemisahan data ke dalam empat kelompok memberikan keseimbangan terbaik antara kompleksitas dan homogenitas cluster.



Gambar 5 Elbow method

Pada Gambar 6 Algoritma *K-Means* berhasil membagi wilayah Jawa Barat ke dalam tiga kelas kerawanan: rendah, sedang dan tinggi. Namun, kelemahan *K-Means* terletak pada ketidakmampuannya menangani *cluster* berdensitas berbeda, bentuk *non-linier*, dan keberadaan *outlier*, yang menjadi penting dalam konteks data bencana.



seragam. Penggabungan ini tidak dilakukan secara paralel, tetapi melalui strategi dua tahap yang saling melengkapi. *Label* akhir dari *hybrid clustering* adalah hasil kombinasi berikut:

- Jika suatu wilayah memiliki *label* yang valid dari *K-Means* dan bukan outlier dari *DBSCAN*, maka label *K-Means* tetap digunakan.
- Jika wilayah diklasifikasikan sebagai *noise* atau *outlier* oleh *DBSCAN*, maka diberi label “ekstrem” atau dipisahkan sebagai kategori khusus.

Dengan pendekatan ini, struktur klasifikasi tetap konsisten secara global namun sensitif terhadap pola lokal yang penting untuk mitigasi bencana. Strategi ini memastikan bahwa keputusan berbasis lokasi dapat lebih akurat dan adaptif, terutama dalam kebijakan penanganan daerah rawan longsor di Jawa Barat. Hasil *Hybrid Clustering* disajikan pada Tabel 3.

Tabel 3 Hasil *hybrid clustering*

No	KABUPATEN	KMeans	DBSCAN	Hybrid	Strategi
1	Kabupaten bogor	2	-1	outlier	evakuasi dini, sistem peringatan cepat,patroli longsor
2	Kabupaten sukabumi	0	-1	outlier	evakuasi dini, sistem peringatan cepat,patroli longsor
3	Kabupaten cianjur	0	-1	outlier	evakuasi dini, sistem peringatan cepat,patroli longsor
4	Kabupaten bandung	0	0	cluster_0	sosialisasi, edukasi rutin kepada masyarakat
5	Kabupaten garut	1	-1	outlier	evakuasi dini, sistem peringatan cepat,patroli longsor
6	Kabupaten tasikmalaya	0	-1	outlier	evakuasi dini, sistem peringatan cepat,patroli longsor

4.5 Validasi Model *Clustering*

Dalam penelitian ini, validasi dilakukan untuk mengevaluasi kualitas hasil pengelompokan data menggunakan metode *K-Means*, *DBSCAN*, dan *hybrid clustering*. Karena *clustering* termasuk metode *unsupervised*, maka validasi dilakukan tanpa label eksternal. Oleh karena itu, digunakan pendekatan evaluasi internal yang bertujuan menilai struktur dan konsistensi antar cluster yang terbentuk.

Pada Tabel 4 terdapat dua metrik utama yang digunakan. Pertama adalah *Silhouette Coefficient*, yaitu ukuran yang menunjukkan sejauh mana suatu data cocok dengan *cluster*-nya sendiri dibandingkan dengan *cluster* lain. Nilai *Silhouette* berkisar antara -1 hingga 1, di mana nilai mendekati 1 menunjukkan bahwa pembentukan cluster dilakukan dengan baik dan jelas terpisah [18].

Metrik kedua adalah *Davies-Bouldin Index (DBI)*. Metrik ini mengukur rata-rata kesamaan antar *cluster* berdasarkan jarak dan sebaran titik data dalam *cluster*. Nilai *DBI* yang lebih rendah menunjukkan pemisahan antar cluster yang lebih jelas dan kualitas pembentukan kelompok yang lebih baik. Dengan menggunakan kedua metrik ini, hasil clustering dapat dinilai secara objektif meskipun tanpa informasi label awal [19].

Tabel 4. Hasil validasi model clustering

Metode	Silhouette Score	Davies-Bouldin Index
K-Means	0.419	0.688
DBSCAN	0.374	0.815
Hybrid (K + DBSCAN)	0.448	0.602

Pada Tabel 4 menunjukkan bahwa:

- Hybrid clustering* memiliki skor *Silhouette* tertinggi (0.448) dan *DBI* terendah (0.602), yang mengindikasikan keseimbangan terbaik antara kepadatan internal dan pemisahan antar cluster.
- K-Means* menunjukkan performa stabil, namun cenderung menyederhanakan kelompok, terutama ketika *outlier* atau bentuk *cluster* tidak beraturan.
- DBSCAN* mampu mengidentifikasi *noise* dan *cluster non-linier*, tetapi nilai validasi internalnya sedikit lebih rendah akibat jumlah cluster bervariasi dan sensitivitas terhadap parameter.

Validasi metrik ini memperkuat bahwa strategi *hybrid clustering* memberikan hasil paling optimal untuk klasifikasi spasial data bencana. Dengan memanfaatkan struktur global dari *K-Means* dan kemampuan deteksi lokal dari *DBSCAN*, model ini mampu menghasilkan klasifikasi yang lebih terstruktur, adaptif, dan representatif terhadap pola kejadian di lapangan.

4.6 Implementasi Web-GIS dan Visualisasi

Sebagai upaya akhir dalam mentransformasikan hasil analisis menjadi alat pendukung pengambilan keputusan, penelitian ini mengembangkan sebuah sistem *Web-GIS* interaktif yang menampilkan hasil klasifikasi wilayah rawan longsor berdasarkan metode *K-Means*, *DBSCAN*, dan *hybrid*. Sistem ini bertujuan untuk memudahkan visualisasi spasial data bencana dan menjadi media interaktif dalam menyampaikan informasi rawan bencana secara geografis.

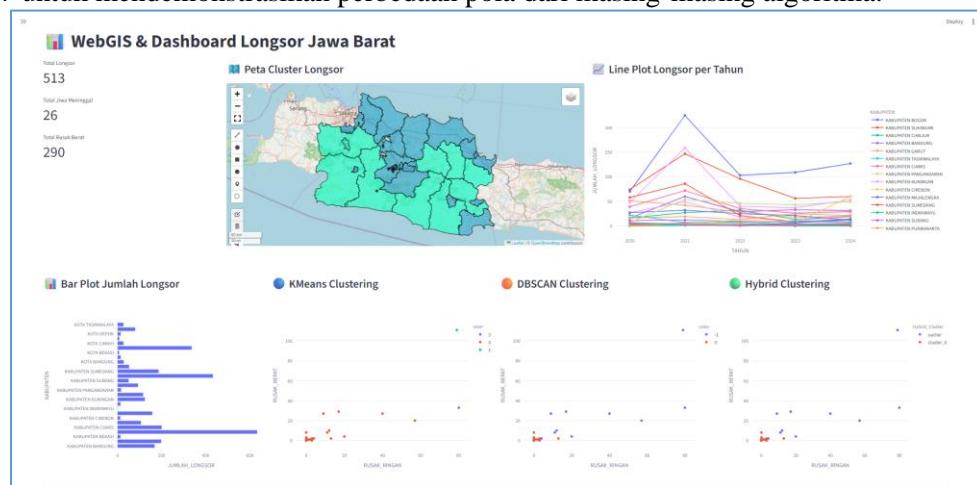
Web-GIS dikembangkan menggunakan *framework Streamlit*, yang memungkinkan integrasi langsung antara hasil analisis *Python* dengan antarmuka *web*. Untuk menampilkan peta spasial, digunakan pustaka *Folium* yang memanfaatkan *leaflet.js* sehingga memungkinkan pengguna berinteraksi langsung dengan peta. *GeoJSON* kabupaten/kota Jawa Barat digunakan sebagai *layer* utama, sementara hasil clustering disimpan dalam format *CSV* dan diolah menjadi *layer* dinamis. Hasil integrasi *clustering* dengan map *geojson* disajikan pada Gambar 9.



Gambar 9 Hasil integrasi clustering dengan map geojson

Penelitian ini tidak hanya menghasilkan model klasifikasi wilayah rawan longsor berbasis *hybrid clustering*, tetapi juga terintegrasi dalam satu *platform* berbasis *dashboard*. Sistem ini berfungsi sebagai jembatan antara hasil analisis ilmiah dengan kebutuhan informasi praktis yang dapat diakses oleh pemangku kepentingan dan masyarakat luas. Halaman utama dari sistem *Web-GIS* menampilkan *dashboard* analitik yang menyajikan gambaran umum status rawan longsor di Jawa Barat melalui beberapa visual utama yang disajikan pada Gambar 10 :

- Peta Provinsi Jawa Barat yang menampilkan klasifikasi rawan longsor berdasarkan pilihan metode (*K-Means*, *DBSCAN*, *Hybrid*) dan tahun (2020–2024).
- Barplot* jumlah kejadian longsor per kabupaten/kota, sebagai gambaran distribusi spasial.
- Scatter Plot Clustering* hasil tiga metode sekaligus, yang menampilkan distribusi dan pemisahan *cluster* untuk mendemonstrasikan perbedaan pola dari masing-masing algoritma.



Gambar 10 Halaman dashboard web gis

Peta interaktif mendukung fungsionalitas *zoom* dan *klik* pada wilayah kabupaten/kota, yang akan menampilkan informasi detail seperti jumlah kejadian, dampak kerusakan, klasifikasi *cluster*, serta strategi mitigasi yang direkomendasikan berdasarkan kategori risikonya. Setiap *cluster* memiliki warna khas dan *label* klasifikasi, termasuk kategori khusus seperti *outlier* hasil *DBSCAN*.

5 Kesimpulan

Penelitian ini berhasil mengembangkan metode *hybrid clustering* dengan menggabungkan algoritma *K-Means* dan *DBSCAN* untuk mengklasifikasikan wilayah rawan longsor di Provinsi Jawa Barat berdasarkan data spasial dan temporal tahun 2020–2024. Hasil menunjukkan bahwa *K-Means* efektif dalam membentuk struktur klasifikasi global berdasarkan karakteristik umum wilayah ditunjukkan dengan *Silhouette Score* sebesar 0.419 dan *Davies-Bouldin Index* sebesar 0.688, sementara *DBSCAN* berperan penting dalam mengidentifikasi outlier dan pola lokal yang tidak terdeteksi oleh pendekatan konvensional ditunjukkan dengan *Silhouette Score* sebesar 0.374 dan *Davies-Bouldin Index* sebesar 0.815. *Hybrid clustering* yang dikembangkan mampu meningkatkan kualitas hasil klasifikasi, terbukti melalui evaluasi menggunakan *Silhouette Score* sebesar 0.448 dan *Davies-Bouldin Index* sebesar 0.602 yang menunjukkan performa lebih baik dibanding metode tunggal. Model *hybrid* ini mampu membagi wilayah ke dalam empat kategori kerawanan rendah, sedang, tinggi, dan sangat tinggi serta menunjukkan dinamika spasial-temporal yang relevan sebagai dasar mitigasi adaptif. Hasil analisis kemudian diimplementasikan dalam bentuk *Web-GIS* interaktif, yang mempermudah visualisasi klasifikasi per tahun dan metode. Secara keseluruhan, pendekatan ini memberikan solusi teknis yang lebih akurat dan aplikatif dalam mendukung pengambilan keputusan berbasis data spasial untuk penanggulangan bencana longsor di Provinsi Jawa Barat.

Referensi

- [1] M. Firman, A. Halik, and L. Septiana, “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*, Vol. 6, No. 4, pp. 856–870, 2022, DOI: 10.52362/jisamar.v6i4.939.
- [2] N. Nurahman and N. A. Tanjung, “*Clustering Village Development in West Java Province on the Condition of Developing Village Strata using K-Means Algorithm*,” *Jurnal Penelitian Pendidikan IPA*, Vol. 9, No. SpecialIssue, pp. 299–306, Dec. 2023, DOI: 10.29303/jppipa.v9ispecialissue.5937.
- [3] A. R. Azzahra, P. Kristalina, and N. Sa’Adah, “*East Java Rivers Characterization using Analytical Hierarchical Clustering on Supporting Disaster Management System*,” in *2024 International Electronics Symposium (IES)*, 2024, pp. 650–655. DOI: 10.1109/IES63037.2024.10665866.
- [4] K. M. P. Dewi, M. L. Tauryawati, and A. F. Zainuddin, “*Clustering of Disaster Prone Areas in Java Island*,” *AIP Conf Proc*, Vol. 3201, No. 1, p. 020003, Nov. 2024, DOI: 10.1063/5.0230700.
- [5] I. S. Fauzi, N. Nuraini, R. W. S. Ayu, and B. W. Lestari, “*Temporal Trend and Spatial Clustering of the Dengue Fever Prevalence in West Java, Indonesia*,” *Heliyon*, Vol. 8, No. 8, p. e10350, 2022, DOI: <https://doi.org/10.1016/j.heliyon.2022.e10350>.
- [6] S. H. Jessen, Z. G. Ma, F. D. Wijaya, J. C. Vasquez, J. Guerrero, and B. N. Jørgensen, “*Identification of Natural Disaster Impacted Electricity Load Profiles with K-Means Clustering Algorithm*,” *Energy Informatics*, Vol. 5, Dec. 2022, DOI: 10.1186/s42162-022-00237-0.
- [7] R. A. Ekaputra, C. Lee, S. H. Kee, and J. J. Yee, “*Emergency Shelter Geospatial Location Optimization for Flood Disaster Condition: A Review*,” Oct. 01, 2022, *MDPI*. DOI: 10.3390/su141912482.
- [8] N. Dwityanti, S. A. Kumala, and S. D. Handayani, “*Comparative Study of Earthquake Clustering in Indonesia using K-Medoids, K-Means, DBSCAN, Fuzzy C-Means and K-AP Algorithms*,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, Vol. 8, No. 6, pp. 768–778, Dec. 2024, DOI: 10.29207/resti.v8i6.5514.

- [9] R. Oktarina and Junita, “Determine the Clustering of Cities in Indonesia for Disaster Management using K-Means by Excel and RapidMiner,” in *IOP Conference Series: Earth and Environmental Science*, IOP Publishing Ltd, Aug. 2021. DOI: 10.1088/1755-1315/794/1/012094.
- [10] M. Shafapourtehrany, P. Yariyan, H. Özener, B. Pradhan, and F. Shabani, “Evaluating the Application of K-Mean Clustering in Earthquake Vulnerability Mapping of Istanbul, Turkey,” *International Journal of Disaster Risk Reduction*, Vol. 79, p. 103154, 2022, DOI: <https://doi.org/10.1016/j.ijdrr.2022.103154>.
- [11] A. Prasetyadi, B. Nugroho, and M. G. Putra, “Determining Natural Disaster Mitigation Level using Unsupervised K-Means Clustering,” in *2022 5th International Conference on Networking, Information Systems and Security: Envisage Intelligent Systems in 5g//6G-based Interconnected Digital Worlds (NISS)*, 2022, pp. 1–5. DOI: 10.1109/NISS55057.2022.10085620.
- [12] T. Handhayani and Z. Rusdi, “K-Means using Dynamic Time Warping for Clustering Cities in Java Island According to Meteorological Conditions,” in *2023 Eighth International Conference on Informatics and Computing (ICIC)*, 2023, pp. 1–6. DOI: 10.1109/ICIC60109.2023.10381899.
- [13] K. E. Setiawan, A. Kurniawan, A. Chowanda, and D. Suhartono, “Clustering Models for Hospitals in Jakarta using Fuzzy C-Means and K-Means,” *Procedia Comput SCI*, Vol. 216, pp. 356–363, 2023, DOI: <https://doi.org/10.1016/j.procs.2022.12.146>.
- [14] S. U. Afifah, A. Fariza, and A. Basofi, “Spatial-Temporal Visualization of Tuberculosis Vulnerability in East Java, Indonesia, using the K-Medoids Clustering Algorithm,” in *2022 International Electronics Symposium (IES)*, 2022, pp. 442–447. DOI: 10.1109/IES55876.2022.9888389.
- [15] W. Hadikurniawati, K. D. Hartomo, and I. Sembiring, “Spatial Clustering of Child Malnutrition in Central Java: A Comparative Analysis using K-Means and DBSCAN,” in *2023 International Conference on Modeling & E-Information Research, Artificial Learning and Digital Applications (ICMERALDA)*, 2023, pp. 242–247. DOI: 10.1109/ICMERALDA60125.2023.10458202.
- [16] E. Choi and J. Song, “Clustering-based Disaster Resilience Assessment of South Korea Communities Building Portfolios using Open GIS and Census Data,” *International Journal of Disaster Risk Reduction*, Vol. 71, p. 102817, 2022, DOI: <https://doi.org/10.1016/j.ijdrr.2022.102817>.
- [17] J. L. Duykers, K. Ardana, R. Yulistiani, E. Irwansyah, and D. Fitrianah, “Identifying Factors for Supporting Early Warning Flood using Clustering Approach and Geo-Spatial Analysis,” *Procedia Comput SCI*, Vol. 227, pp. 540–547, 2023, DOI: <https://doi.org/10.1016/j.procs.2023.10.556>.
- [18] K. H. Izzuddin and A. W. Wijayanto, “Pemodelan Clustering Ward, K-Means, DIANA, dan PAM dengan PCA untuk Karakterisasi Kemiskinan Indonesia Tahun 2021,” *Jurnal Sistem Komputer*, Vol. 13, No. 1, p. 2020, 2024, DOI: 10.34010/komputika.v13i1.10803.
- [19] S. Sun, K. Lei, Z. Xu, W. Jing, and G. Sun, “Analysis of K-Means and K-DBSCAN Commonly used in Data Mining,” in *2023 International Conference on Intelligent Media, Big Data and Knowledge Mining (IMBDKM)*, 2023, pp. 37–41. DOI: 10.1109/IMBDKM57416.2023.00014.