

Implementasi Algoritma *K-Means* untuk *Clustering Zonasi Daerah Rawan Bencana Alam* di Kabupaten Pati

Implementation of the K-Means Algorithm for Clustering Zoning of Natural Disaster-Prone Areas in Pati Regency

¹Arifin*, ²Bambang Agus Herlambang, ³Aris Tri Joko Harjanto

^{1,2,3}Program Studi Informatika, Fakultas Teknik dan Informatika, Universitas PGRI Semarang
^{1,2,3}Jl. Sidodadi Timur No.24, Karangtempel, Kec. Semarang Tim., Kota Semarang, Jawa Tengah
Indonesia

*e-mail: carifeen@upgris.ac.id

(received: 19 December 2025, revised: 12 January 2026, accepted: 16 January 2026)

Abstrak

Berdasarkan data Badan Pusat Statistik (BPS) Kabupaten Pati, selama periode 2018-2022 Kabupaten Pati sering dilanda bencana alam banjir dan tanah longsor di sebagian banyak wilayah. Banyaknya bencana alam yang melanda dan kurangnya pemetaan terhadap daerah yang mengalami bencana merupakan kendala yang perlu diselesaikan dengan tepat. Memanfaatkan teknologi dan metode *K-Means*, penelitian ini memberikan alternatif solusi yang efektif dalam rangka mengidentifikasi dan memetakan daerah yang rentan terhadap bencana alam. Hasil analisis dari penelitian yang telah diselesaikan menunjukkan bahwa Kabupaten Pati dibagi menjadi 3 *cluster*: *cluster* 1 untuk daerah sangat rawan bencana 42.86% (9 wilayah), *cluster* 2 untuk daerah rawan bencana 19.04% (4 wilayah) dan *cluster* 3 untuk daerah tidak rawan bencana 38.1% (8 wilayah). Hasil visualisasi merupakan peta klasifikasi wilayah berdasarkan tingkat kerawanan bencana yang dapat dijadikan dasar bagi pemerintah daerah maupun instansi terkait dalam merumuskan kebijakan penanggulangan bencana yang lebih tepat sasaran.

Kata kunci: penanggulangan bencana, banjir, , k-means, tanah longsor

Abstract

Based on data from the Central Bureau of Statistics (BPS) of Pati Regency, during the 2018–2022 period the region frequently experienced natural disasters, particularly floods and landslides, across many areas. The high frequency of these disasters and the lack of proper mapping of affected areas present challenges that need to be addressed effectively. By utilizing technology and the *K-Means* clustering method, this study proposes an alternative solution to identify and map areas that are vulnerable to natural disasters. The results of the analysis indicate that Pati Regency can be divided into three clusters: Cluster 1 represents highly disaster-prone areas, accounting for 42.86% (9 regions); Cluster 2 represents disaster-prone areas, accounting for 19.04% (4 regions); and Cluster 3 represents areas with low disaster vulnerability, accounting for 38.1% (8 regions). The visualization results are presented as a classification map of regions based on their disaster vulnerability levels. This map can serve as a reference for local governments and relevant institutions in formulating more targeted and effective disaster mitigation policies.

Keywords: disaster management, flood, k-means, landslide

1 Pendahuluan

Indonesia merupakan salah satu negara dengan tingkat kerentanan bencana alam tertinggi di dunia, terletak pada pertemuan empat lempeng tektonik utama: Lempeng Pasifik, Eurasia, Indo-Australia, dan Filipina [1]. Kabupaten Pati merupakan salah satu wilayah kabupaten yang terletak di Provinsi Jawa Tengah, Indonesia, dengan letak geografis yang cukup strategis di bagian pesisir utara Pulau Jawa. Secara topografis, wilayah Kabupaten Pati terdiri dari dataran rendah yang berada di bagian utara dan tengah, serta bagian perbukitan dan pegunungan di bagian selatan yang termasuk dalam kawasan Pegunungan Kendeng [2]. Kondisi ini menjadikan wilayah kabupaten Pati memiliki

<http://sistemasi.ftik.unisi.ac.id>

keragaman karakteristik geospasial yang signifikan terhadap berbagai aspek pembangunan dan kerentanan terhadap bencana, terutama banjir dan tanah longsor [3].

Pati merupakan salah satu kota dimana hampir setiap tahunnya dilanda bencana banjir [4]. Bencana banjir merupakan aspek interaksi manusia dengan alam yang timbul dari proses dimana manusia mencoba menggunakan alam yang bermanfaat dan menghindari alam yang merugikan manusia [5]. Kemudian, alih fungsi lahan pertanian menjadi perumahan atau infrastruktur lainnya dapat mengganggu keseimbangan ekosistem lokal [6],[7]. Peningkatan frekuensi dan intensitas banjir tidak hanya berdampak pada kerusakan fisik, tetapi juga menimbulkan kerugian ekonomi bagi masyarakat. Kerusakan infrastruktur seperti jalan raya dan jalan setapak mengganggu mobilitas dan aksesibilitas, yang pada gilirannya mempengaruhi aktivitas ekonomi dan sosial[7]. Berdasarkan data yang dihimpun secara online melalui website Badan Pusat Statistik (BPS) Kabupaten Pati, selama periode 2018-2022 Kabupaten Pati sering dilanda bencana alam banjir dan tanah longsor di sebagian banyak wilayah seperti kecamatan Sukolilo, Kayen, Juwana, Jakenan, Gabus dan yang lainnya (<https://patikab.bps.go.id/>).

Berbagai sistem informasi dan komunikasi sangat penting untuk persiapan dan respons bencana [8]. Penanganan dan mitigasi risiko bencana memerlukan langkah yang sistematis dan berbasis data. Salah satu upaya penting dalam manajemen bencana adalah pemetaan zonasi daerah rawan bencana. Pemetaan risiko bencana adalah kegiatan pembuatan peta yang merepresentasikan dampak negatif yang dapat timbul berupa kerugian materi dan non materi pada suatu wilayah apabila terjadi bencana. Diperlukan data yang valid untuk proses pemetaan risiko sehingga dapat mempresentasikan kondisi sebenarnya dilapangan [9].

Urgensi dalam penelitian *clustering* zonasi untuk memetakan daerah yang berpotensi rawan bencana. Selanjutnya pemerintah daerah mampu memberikan edukasi kepada masyarakat yang tinggal di wilayah-wilayah tersebut. Metode yang dapat digunakan dalam pengelompokan data adalah algoritma K-Means *Clustering* dengan pendekatan partisi (*centroid*) karena mengelompokkan data kedalam sejumlah klaster yang ditentukan. Algoritma ini mampu mengelompokkan wilayah berdasarkan kesamaan karakteristik tertentu sehingga dapat membantu dalam membentuk pola zonasi rawan bencana yang lebih objektif [10].

Penelitian ini bertujuan untuk melakukan *clustering* atau pengelompokan wilayah rawan bencana banjir dan tanah longsor di Kabupaten Pati menggunakan metode K-Means serta mengembangkan sistem informasi geografis (SIG) berbasis web yang mampu memodelkan klastering zonasi bencana banjir dan tanah longsor di Kabupaten Pati. Dengan pendekatan ini, diharapkan dapat diperoleh peta klasifikasi wilayah berdasarkan tingkat kerawanan bencana yang dapat dijadikan dasar bagi pemerintah daerah maupun instansi terkait dalam merumuskan kebijakan penanggulangan bencana yang lebih tepat sasaran.

2 Tinjauan Literatur

Berbagai penelitian terdahulu menunjukkan bahwa algoritma K-Means merupakan metode yang efektif dalam melakukan klasterisasi wilayah rawan bencana alam. Penerapan algoritma ini telah dilakukan pada berbagai skala wilayah dan dengan beragam pendekatan data. Penelitian di Kabupaten Mandailing Natal membuktikan bahwa algoritma K-Means mampu mengelompokkan daerah rawan bencana banjir melalui tahapan normalisasi data, penentuan jumlah *cluster*, inisialisasi centroid, serta proses iterasi hingga diperoleh hasil klasterisasi yang optimal. Hasil pengelompokan tersebut kemudian diimplementasikan dalam sebuah sistem berbasis web menggunakan bahasa pemrograman PHP, sehingga memudahkan dalam pengelolaan dan penyajian informasi kerawanan bencana [11].

Pendekatan serupa juga diterapkan pada wilayah yang lebih luas, seperti penelitian di Provinsi Jawa Tengah, yang menggunakan algoritma K-Means untuk membentuk wilayah ke dalam beberapa *cluster* berdasarkan tingkat kerawanan bencana. Hasil klasterisasi divisualisasikan menggunakan library *geopandas*, yang memungkinkan penyajian data secara spasial sehingga pola persebaran daerah rawan bencana dapat diamati dengan lebih jelas [12].

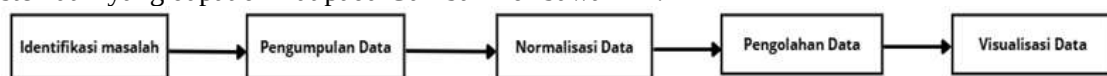
Selain itu, penelitian lain yang memanfaatkan data bencana alam berdasarkan aspek wilayah, waktu kejadian, jumlah korban, serta tingkat kerusakan fasilitas menunjukkan bahwa algoritma K-Means mampu mengelompokkan wilayah ke dalam beberapa *cluster* berdasarkan kriteria tersebut. Pendekatan ini memberikan pemahaman yang lebih terperinci mengenai pola dan distribusi bencana

alam, serta menunjukkan bahwa setiap wilayah dapat memiliki karakteristik kerawanan yang berbeda. Penelitian tersebut juga menegaskan bahwa sebagian besar wilayah dapat dikelompokkan secara konsisten, kecuali pada pengelompokan yang didasarkan pada periode waktu tertentu seperti bulanan [13].

Berdasarkan uraian penelitian terdahulu tersebut, dapat disimpulkan bahwa algoritma K-Means memiliki fleksibilitas dan keandalan dalam mengelompokkan wilayah rawan bencana alam, baik berdasarkan dampak, frekuensi, maupun karakteristik wilayah. Namun demikian, sebagian besar penelitian masih berfokus pada pengelompokan wilayah berdasarkan dampak bencana secara umum, sehingga penelitian lanjutan masih diperlukan untuk mengintegrasikan hasil klasterisasi dengan analisis zonasi kerawanan bencana pada tingkat wilayah yang lebih spesifik, seperti kabupaten, guna mendukung perencanaan mitigasi bencana yang lebih tepat sasaran. Perbedaan dengan penelitian ini yaitu kebaharuan dalam dalam proporsi dan normalisasi data untuk menginisiasi *centroid* sehingga *cluster* yang didapatkan akan lebih berkualitas dan hasil analisis mitigasi dapat digunakan sebagai dasar penanggulangan bencana yang lebih tepat sasaran.

3 Metode Penelitian

Tahapan pada penelitian ini adalah langkah-langkah sistematis yang dilakukan oleh peneliti guna memperoleh data serta informasi yang dibutuhkan. Pada penelitian ini juga, peneliti menggunakan kerangka kerja yang berfungsi sebagai acuan supaya penelitian dapat dilakukan secara terorganisir dan sistematis yang dapat dilihat pada Gambar 1 di bawah ini:



Gambar 1 Alur penelitian

1. Clustering

Clustering merupakan Proses yang di gunakan untuk melibatkan perancangan anggota sehingga seluruh anggota setiap kelompok mempunyai kesamaan terhadap matriks tertentu [14].

2. Metode Clustering K-Means

K-means merupakan salah satu algoritma dalam machine learning yang digunakan untuk mengelompokkan data menjadi beberapa cluster atau kelompok berdasarkan karakteristik yang serupa. Dalam konteks K-Means, huruf K mewakili jumlah kelompok (*cluster*) yang akan ditetapkan. Pada tahap berikutnya, nilai K ditentukan secara acak untuk keperluan inisialisasi. Adapun “means” adalah nilai tengah yang berperan sebagai pusat dari masing-masing klaster atau yang dikenal sebagai *centroid* [15]. Prosedur pelaksanaan Metode K-Means dijabarkan sebagai berikut :

- Menetapkan banyaknya cluster (*k*) yang akan digunakan pada data.
- Menetapkan centroid awal dengan cara acak.
- Perhitungan jarak dilakukan dengan menentukan jarak terdekat ke centroid dengan menerapkan rumus (1) berikut ini:

$$d = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \quad (1)$$

Dimana:

d = Euclidean Distance

x, y = Koordinat Posisi objek

s, t = Koordinat Posisi *centroid*

i = Banyak objek

- Objek dikelompokkan sesuai dengan jarak terdekat ke centroid.
- Menghitung nilai rata-rata pada setiap kelompok dengan menggunakan rumus (2) berikut:

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj} \quad (2)$$

Keterangan:

V_{ij} = Centroid rata-rata cluster ke-*i* untuk variable ke-*j*.

N_i = Jumlah anggota cluster ke-*i*.

- i, k = Indeks dari cluster.
 j = Indeks dari variable.
 X_{kj} = Nilai data ke- k variable ke- j untuk cluster tersebut.

f. Mengulangi langkah pada poin 3 dan poin 4 secara iteratif hingga diperoleh centroid dengan nilai yang terbaik [16].

3. Sistem Informasi Geografis berbasis WEB

Sistem informasi geografis berbasis web atau juga disebut dengan WebGIS atau InternetGIS, didefinisikan sebagai suatu jaringan (network) berbasis layanan informasi geografis yang memanfaatkan internet baik menggunakan jaringan kabel (wired) maupun tanpa kabel (wireless) untuk mengakses informasi geografis maupun sebagai tools guna melakukan *spatial* analisis [17].

4. PHP

Merupakan bahasa pelengkap HTML yang memungkinkan dibuatnya aplikasi dinamis yang memungkinkan adanya pengolahan data dan pemrosesan data. Semua syntax yang diberikan akan sepenuhnya dijalankan pada server sedangkan yang dikirimkan ke browser hanya hasilnya saja. Kemudian merupakan bahasa berbentuk script yang ditempatkan dalam server dan diproses di server. Hasilnya akan dikirimkan ke *client*, tempat pemakai menggunakan browser [18].

4 Hasil dan Pembahasan

Pengujian perhitungan algoritma K-Means dilakukan dengan metode pengujian perhitungan secara manual. Pengujian di luar sistem ini bertujuan untuk mendapatkan hasil klasterisasi wilayah yang akan diimplementasikan kedalam pemetaan sistem informasi geografis (SIG). Pengujian ini menggunakan data tentang daerah yang mengalami bencana alam banjir dan tanah longsor menurut kecamatan di Kabupaten Pati yang diperoleh dari Web BPS Pati. Pada tahap clustering, data dihitung berdasarkan jumlah kejadian daerah rawan bencana selama periode 2018 – 2022 pada 21 kecamatan di Kabupaten Pati. Rentang waktu lima tahun dipilih untuk mewakili kondisi kerawanan bencana dalam jangka menengah, sehingga pola kejadian bencana yang dianalisis tidak hanya bersifat sesaat, tetapi mencerminkan kecenderungan yang relatif stabil di setiap wilayah. Periode 2018 – 2022 juga mencakup variasi intensitas bencana, termasuk tahun-tahun dengan kejadian bencana yang cukup signifikan, sehingga mampu memberikan gambaran yang lebih komprehensif mengenai tingkat kerawanan. Selain itu data setelah tahun 2022 belum seluruhnya tersedia atau belum dipublikasikan secara lengkap oleh BPS pada saat penelitian dilakukan. Oleh karena itu, penggunaan data hingga tahun terakhir berpotensi menimbulkan ketidakkonsistenan dan mengurangi validitas hasil analisis.

Pada Tabel 1 dan Tabel 2 merupakan data awal sebelum dilakukan penyesuaian proporsi dan normalisasi.

Tabel 1 Data jumlah desa terkena banjir pertahun

No	WILAYAH	DESA	2018	2019	2020	2021	2022
1	010. Sukolilo	16	3	7	3	3	4
2	020. Kayen	17	9	3	4	3	6
3	030. Tambakromo	18	2	0	2	0	9
4	040. Winong	30	3	0	0	0	0
5	050. Pucakwangi	20	3	0	0	0	0
6	060. Jaken	21	12	5	0	0	1
7	070. Batangan	18	9	1	0	0	7
8	080. Juwana	29	15	2	0	11	6
9	090. Jakenan	23	10	5	4	8	6
10	100. Pati	29	18	0	0	5	4
11	110. Gabus	24	6	3	2	11	19
12	120. Margorejo	18	1	3	1	0	2
13	130. Gembong	11	1	0	0	0	0
14	140. Tlogowungu	15	0	0	0	0	0
15	150. Wedarijaksa	18	2	2	2	0	3
16	160. Trangkil	16	1	1	0	0	2
17	170. Margoyoso	22	6	3	0	0	3
18	180. Gunungwungkal	15	0	0	0	0	0
19	190. Cluwak	13	0	0	0	0	0
20	200. Tayu	21	2	5	2	0	0
21	210. Dukuhseti	12	8	2	3	0	1

Tabel 2 Data jumlah desa terkena tanah longsor pertahun

No	Wilayah	Desa	2018	2019	2020	2021	2022
1	010. Sukolilo	16	0	0	2	0	1
2	020. Kayen	17	1	0	3	0	0
3	030. Tambakromo	18	1	0	1	1	0
4	040. Winong	30	1	1	0	0	0
5	050. Pucakwangi	20	2	0	0	0	0
6	060. Jaken	21	1	0	0	0	0
7	070. Batangan	18	0	0	0	0	0
8	080. Juwana	29	0	0	0	0	0
9	090. Jakenan	23	0	0	0	0	0
10	100. Pati	29	0	0	0	0	1
11	110. Gabus	24	0	0	0	0	0
12	120. Margorejo	18	0	1	0	0	0
13	130. Gembong	11	3	0	0	1	0
14	140. Tlogowungu	15	1	1	0	2	0
15	150. Wedarijaksa	18	0	0	0	0	0
16	160. Trangkil	16	0	0	0	0	0
17	170. Margoyoso	22	0	0	0	0	0
18	180. Gunungwungkal	15	5	0	0	1	1
19	190. Cluwak	13	1	0	0	0	0
20	200. Tayu	21	0	0	0	0	0
21	210. Dukuhseti	12	0	0	0	0	0

Sebelum menyelesaikan perhitungan Algoritma K-Means diperlukan normalisasi jumlah desa terdampak bencana terhadap total desa per kecamatan dilakukan dalam bentuk proporsi, sebagaimana umum digunakan dalam analisis statistik deskriptif [19]. Dalam statistik, menghitung proporsi dengan menggunakan rumus (3) berikut ini:

$$Proporsi = \frac{Bagian (Subset)}{Total(keseluruhan)} \quad (3)$$

Tabel 3 Hasil proporsi desa terkena banjir pertahun

No	WILAYAH	2018	2019	2020	2021	2022
1	010. Sukolilo	0.19	0.44	0.19	0.19	0.25
2	020. Kayen	0.53	0.17	0.23	0.17	0.35
3	030. Tambakromo	0.11	0	0.11	0	0.5
4	040. Winong	0.1	0	0	0	0
5	050. Pucakwangi	0.15	0	0	0	0
6	060. Jaken	0.57	0.24	0	0	0.05
7	070. Batangan	0.5	0.06	0	0	0.39
8	080. Juwana	0.52	0.07	0	0.38	0.21
9	090. Jakenan	0.43	0.22	0.17	0.35	0.26
10	100. Pati	0.62	0	0	0.17	0.14
11	110. Gabus	0.25	0.13	0.08	0.46	0.79
12	120. Margorejo	0.06	0.17	0.06	0	0.11
13	130. Gembong	0.09	0	0	0	0
14	140. Tlogowungu	0	0	0	0	0
15	150. Wedarijaksa	0.11	0.11	0.11	0	0.17
16	160. Trangkil	0.06	0.06	0	0	0.13
17	170. Margoyoso	0.27	0.14	0	0	0.14
18	180. Gunungwungkal	0	0	0	0	0
19	190. Cluwak	0	0	0	0	0
20	200. Tayu	0.09	0.24	0.09	0	0
21	210. Dukuhseti	0.67	0.17	0.25	0	0.08

Berdasarkan Tabel 3, terlihat bahwa proporsi desa yang terdampak banjir di Kabupaten Pati pada periode 2018–2022 menunjukkan pola yang berfluktuasi antarwilayah dan antarwaktu. Beberapa kecamatan mengalami tingkat proporsi terdampak banjir yang relatif tinggi pada tahun-tahun tertentu, sementara kecamatan lainnya cenderung menunjukkan nilai yang rendah atau bahkan tidak mengalami kejadian banjir sama sekali.

Tabel 4 Hasil proporsi desa terkena tanah longsor pertahun

No	Wilayah	2018	2019	2020	2021	2022
1	010. Sukolilo	0	0	0.13	0	0.06
2	020. Kayen	0.06	0	0.18	0	0
3	030. Tambakromo	0.05	0	0.05	0.06	0
4	040. Winong	0.03	0.03	0	0	0
5	050. Pucakwangi	0.1	0	0	0	0

<http://sistemasi.ftik.unisi.ac.id>

6	060. Jaken	0.05	0	0	0	0
7	070. Batangan	0	0	0	0	0
8	080. Juwana	0	0	0	0	0
9	090. Jakenan	0	0	0	0	0
10	100. Pati	0	0	0	0	0.03
11	110. Gabus	0	0	0	0	0
12	120. Margorejo	0	0.06	0	0	0
13	130. Gembong	0.27	0	0	0.09	0
14	140. Tlogowungu	0.07	0.07	0	0.13	0
15	150. Wedarijaksa	0	0	0	0	0
16	160. Trangkil	0	0	0	0	0
17	170. Margoyoso	0	0	0	0	0
18	180. Gunungwungkal	0.33	0	0	0.07	0.07
19	190. Cluwak	0.08	0	0	0	0
20	200. Tayu	0	0	0	0	0
21	210. Dukuhseti	0	0	0	0	0

Berdasarkan Tabel 4, proporsi desa yang terdampak tanah longsor di Kabupaten Pati selama periode 2018–2022 menunjukkan intensitas kejadian yang relatif lebih rendah dan tidak merata dibandingkan dengan bencana banjir. Kejadian tanah longsor hanya terjadi di beberapa kecamatan tertentu dan pada tahun-tahun tertentu, sedangkan sebagian besar kecamatan tidak mengalami kejadian tanah longsor.

Setelah proses proporsi data selesai dilakukan, tahapan selanjutnya yang perlu dilakukan adalah normalisasi data guna menyelesaikan algoritma K-Means menggunakan rumus (4) berikut ini [20].

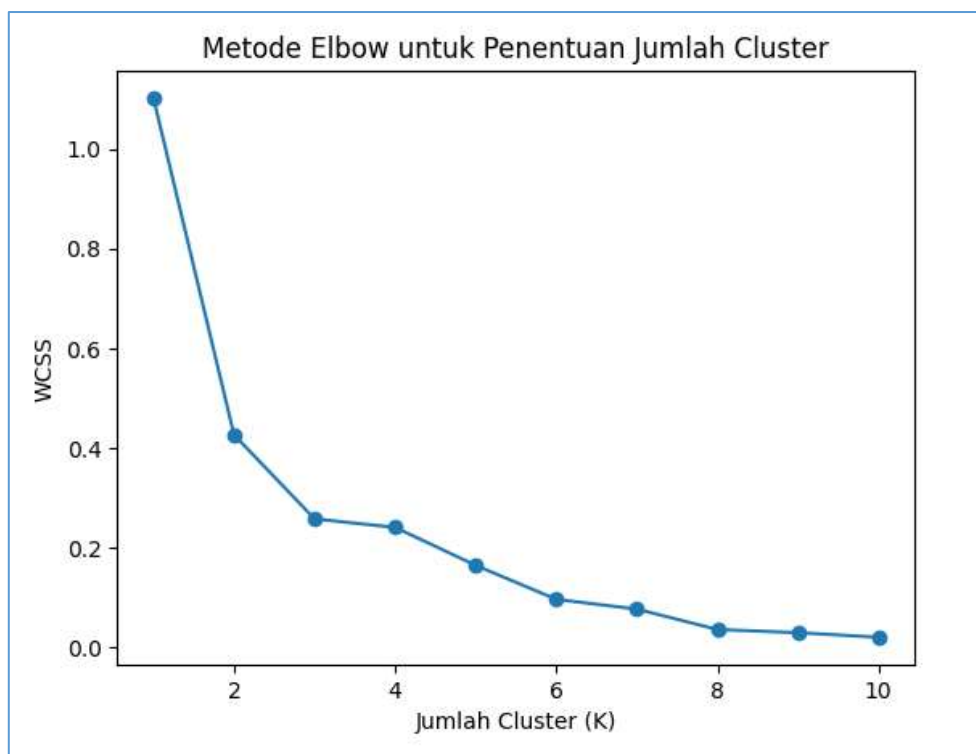
$$Data\ kriteria = \frac{Data\ awal - Min(DataKriteria)}{Max(DataKriteria) - Min(DataKriteria)} \quad (4)$$

Tabel 5 Data hasil normalisasi bencana

No	Wilayah	Banjir	Tanah Longsor
1	010. Sukolilo	0.25	0.3
2	020. Kayen	0.33	0.27
3	030. Tambakromo	0.29	0.6
4	040. Winong	0.2	0.4
5	050. Pucakwangi	0.2	0.2
6	060. Jaken	0.3	0.2
7	070. Batangan	0.38	0
8	080. Juwana	0.45	0
9	090. Jakenan	0.43	0
10	100. Pati	0.3	0.2
11	110. Gabus	0.36	0
12	120. Margorejo	0.47	0.2
13	130. Gembong	0.2	0.27
14	140. Tlogowungu	0	0.4
15	150. Wedarijaksa	0.6	0
16	160. Trangkil	0.4	0
17	170. Margoyoso	0.4	0
18	180. Gunungwungkal	0	0.28
19	190. Cluwak	0	0.2
20	200. Tayu	0.36	0
21	210. Dukuhseti	0.35	0

Berdasarkan Tabel 5, hasil normalisasi data bencana banjir dan tanah longsor menunjukkan variasi tingkat keterpaparan antar wilayah di Kabupaten Pati. Normalisasi ini bertujuan untuk menyamakan skala data dari kedua jenis bencana sehingga dapat digunakan secara simultan dalam proses analisis kluster menggunakan metode K-Means.

Setelah normalisasi data selesai dilakukan, tahap selanjutnya adalah pengelompokan menggunakan algoritma K-Means. Pada proses tahap awal yaitu menentukan jumlah *cluster* untuk kategori yang akan di gunakan untuk analisis data. Metode elbow digunakan untuk menentukan jumlah kluster yang paling optimal seperti yang di tunjukan pada Gambar 2.



Gambar 2 Grafik metode elbow

Penentuan jumlah *cluster* dilakukan menggunakan metode Elbow dengan melihat nilai *Within Cluster Sum of Squares* (WCSS). Berdasarkan grafik Elbow, terlihat bahwa penurunan WCSS sangat signifikan hingga $K = 3$, namun setelah nilai tersebut penurunannya relatif kecil dan cenderung landai. Oleh karena itu, dapat disimpulkan bahwa jumlah *cluster* yang optimal dalam penelitian ini adalah tiga *cluster*.

Tabel 6 Jumlah *cluster*

Cluster	Potensi
C1	Sangat rawan
C2	Rawan
C3	Tidak rawan

Tabel 6 menunjukkan 3 *cluster* yang akan ditetapkan yaitu *cluster* sangat rawan (C1), *cluster* rawan (C2), *cluster* tidak rawan (C3).

Pada penelitian ini penentuan nilai titik *cluster* centroid awal dipilih secara acak dengan mengambil tiga data dengan memperhatikan jumlah daerah yang terdampak bencana alam. Data yang akan digunakan sebagai contoh data centroid awal secara acak dari data kecamatan yang sudah dinormalisasi seperti yang ditunjukkan pada Tabel 7.

Tabel 7 Centroid data iterasi awal

Cluster Potensi	Kecamatan	Daerah Banjir (X)	Daerah Tanah Longsor (Y)
C1	150. Wedarijaksa	0.6	0
C2	120. Margorejo	0.466667	0.2
C3	190. Cluwak	0	0.2

Langkah selanjutnya adalah proses melakukan perhitungan untuk menentukan jarak *Euclidean* dengan masing-masing *centroid* menggunakan persamaan (1) dan proses perhitungan diulang secara iterative sampai dengan data yang diperoleh dari hasil *clustering* ini sama dengan iterasi sebelumnya. Proses perhitungan dimulai dari iterasi awal

$$C1(010. Sukolilo) = \sqrt{(0.25 - 0.6)^2 + (0.3 - 0)^2} = \sqrt{(-0.35)^2 + (0.3)^2} = 0.460977$$

$$C2(010. Sukolilo) = \sqrt{(0.25 - 0.466667)^2 + (0.3 - 0.2)^2} = \sqrt{(-0.216667)^2 + (0.1)^2} = 0.238630$$

$$C3(010. Sukolilo) = \sqrt{(0.25 - 0)^2 + (0.3 - 0.2)^2} = \sqrt{(0.25)^2 + (0.1)^2} = 0.269258$$

Tabel 8 Data hasil clustering iterasi 1

No	Wilayah	X	Y	Jarak ke C1	Jarak ke C2	Jarak ke C3	Cluster	Potensi
1	010. Sukolilo	0.25	0.3	0.460977	0.23863	0.269258	C2	Rawan
2	020. Kayen	0.333333	0.266667	0.377124	0.149071	0.339935	C2	Rawan
...	...	...	...	...	...	...	...	...
19	190. Cluwak	0	0.2	0.632456	0.466667	0	C3	Tidak Rawan
20	200. Tayu	0.36	0	0.24	0.226667	0.411825	C2	Rawan
21	210. Dukuhseti	0.35	0	0.25	0.231541	0.403113	C2	Rawan

Berdasarkan hasil perhitungan iterasi 1 pada Tabel 8, didapatkan data *cluster* dengan 5 wilayah yang masuk kedalam kategori sangat rawan bencana (C1) meliputi kecamatan Juwana, Jakenan, Wedarijaksa, Trangkil, Margoyoso. Pada kategori daerah rawan (C2) bencana terdapat 10 wilayah yaitu kecamatan Sukolilo, Kayen, Tambakromo, Jaken, Batangan, Pati, Gabus, Margorejo, Tayu, dan Dukuhseti. Sedangkan pada kategori daerah tidak rawan bencana (C3) terdapat 6 wilayah yaitu Kecamatan Winong, Gembong, Tlogowungu, Gunungwungkal dan Cluwak.

Setelah hasil perhitungan pada iterasi 1 diperoleh, selanjutnya untuk melakukan perhitungan iterasi 2 dibutuhkan data *centroid* baru menggunakan persamaan (5) berikut ini. Sehingga diperoleh *centroid* baru pada Tabel 9.

$$\text{Centroid baru} = \frac{\text{Nilai rata-rata cluster 1}}{\text{Jumlah data cluster 1}} \quad (5)$$

Cluster 1

$$X = (0.453333+0.453333+0.6+0.4+0.4) : 5 = 0.457333$$

$$Y = (0+0+0+0+0) : 5 = 0$$

Cluster 2

$$X = (0.5+0.333333+0.288889+0.3+0.377778+0.3+0.364706+0.466667+0.36+0.35) : 10 = 0.339137$$

$$Y = (0.3+0.266667+0.6+0.2+0+0.2+0+0.2+0+0) : 10 = 0.176667$$

Cluster 3

$$X = (0.2+0.2+0.2+0+0+0) : 6 = 0.1$$

$$Y = (0.2 + 0.26667 + 0.4 + 0.28 + 0.2) : 6 = 0.291111$$

Tabel 9 Centroid data iterasi 2

Cluster	X	Y
C1	0.457333	0
C2	0.339137	0.176667
C3	0.1	0.291111

Hasil perhitungan pada Tabel 10 menggunakan *centroid* baru pada iterasi ke 2 membentuk *cluster* yang berbeda dengan perhitungan pada itersi awal dimana pada perhitungan iterasi ke dua terdapat 9 wilayah kecamatan masuk pada kategori sangat rawan (C1), 4 wilayah kecamatan masuk kategori rawan (C2) dan 8 wilayah kecamatan masuk dalam kategori tidak rawan (C3), sehingga proses perhitungan jarak perlu dilakukan sampai memperoleh hasil yang sama dengan iterasi sebelumnya.

Tabel 10 Hasil clustering iterasi 2

No	Wilayah	X	Y	Jarak ke C1	Jarak ke C2	Jarak ke C3	Cluster	Potensi
1	Kec. Sukolilo	0.25	0.3	0.3647	0.1522	0.1503	C3	Tidak Rawan
2	Kec. Kayen	0.333	0.2667	0.2941	0.0902	0.2346	C2	Rawan
3	Kec. Tambakromo	0.2889	0.6	0.6232	0.4263	0.3621	C3	Tidak Rawan
...	...	...	...	...	...	...	...	...
19	Kec. Cluwak	0	0.2	0.4992	0.3399	0.1353	C3	Tidak Rawan
20	Kec. Tayu	0.36	0	0.0973	0.1779	0.3903	C1	Sangat rawan
21	Kec. Dukuhseti	0.35	0	0.1073	0.177	0.3837	C1	Sangat rawan

Setelah dilakukan pengelompokan maka *cluster* berdasarkan data yang terdekat dengan *centroid* seperti yang ditunjukkan pada Tabel 11 adalah sebagai berikut.

Tabel 11 Pengelompokan hasil clustering iterasi 1 dan 2

NO	DATA WILAYAH																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Cluster Iterasi 1	C2	C2	C2	C3	C3	C2	C2	C1	C1	C2	C2	C2	C3	C3	C1	C1	C1	C3	C3	C2	C2
Cluster Iterasi 2	C3	C2	C3	C3	C3	C2	C1	C1	C1	C2	C1	C2	C3	C3	C1	C1	C1	C3	C3	C1	C1

Selanjutnya melakukan perhitungan iterasi 3 dengan *centroid* baru menggunakan persamaan (5) sehingga diperoleh *centroid* baru pada Tabel 12.

Tabel 12 Centroid data iterasi 3

Cluster	X	Y
C1	0.415456	0
C2	0.349925	0.216675
C3	0.142363	0.330838

Hasil perhitungan menggunakan *centroid* baru pada iterasi ke 3 pada Tabel 13 membentuk *cluster* yang sama pada perhitungan iterasi ke 2 dimana pada kategori sangat rawan terdapat 9 wilayah yaitu Kecamatan Batangan, Juwana, Jakenan, Gabus, Wedarijaksa, Trangkil, Margoyoso, Tayu dan Dukuhseti. Pada kategori daerah rawan bencana terdapat 4 wilayah yaitu Kecamatan Kayen, Jaken dan Pati, Margorejo. Pada daerah dengan kategori tidak rawan bencana terdapat 8 wilayah meliputi kecamatan Tambakromo, Sukolilo, Winong, Gembong, Pucakwangi, Gunungwungkal, Tlogowungu dan Cluwak sehingga proses perhitungan jarak *Euclidean* sudah dapat dihentikan pada perhitungan iterasi ke 3. Data lengkap hasil perhitungan jarak dapat dilihat pada Tabel 13.

Tabel 13 Hasil clustering iterasi 3

No	Wilayah	X	Y	Jarak ke C1	Jarak ke C2	Jarak ke C3	Cluster	Potensi
1	010. Sukolilo	0.25	0.3	0.3476	0.1271	0.1122	C3	Tidak Rawan
2	020. Kayen	0.333	0.2667	0.2791	0.0528	0.2009	C2	Rawan
3	030. Tambakromo	0.2889	0.6	0.6132	0.3882	0.3065	C3	Tidak Rawan
...	...	...	...	...	...	...	...	...
19	190. Cluwak	0	0.2	0.4612	0.3512	0.1892	C3	Tidak Rawan
20	200. Tayu	0.36	0	0.0555	0.2169	0.3884	C1	Sangat Rawan
21	210. Dukuhseti	0.35	0	0.0655	0.2167	0.3881	C1	Sangat Rawan

Setelah dilakukan pengelompokan berdasarkan pada hasil iterasi 2 dan iterasi 3 maka *cluster* berdasarkan data yang terdekat dengan *centroid* seperti yang ditunjukkan pada Tabel 14 sebagai berikut.

Tabel 14 Pengelompokan hasil clustering iterasi 2 dan 3

NO	DATA WILAYAH																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Cluster Iterasi 2	C3	C2	C3	C3	C3	C2	C1	C1	C1	C2	C1	C	C3	C3	C	C1	C	C	C	C	C
Cluster Iterasi 3	C3	C2	C3	C3	C3	C2	C1	C1	C1	C2	C1	2	C	C3	C	C1	1	3	3	1	1

Untuk memastikan tidak ada perubahan anggota *cluster* maka perlu dilakukan uji validitas stabilitas dengan cara menjalankan algoritma secara berulang menggunakan *centroid* iterasi 3 dst. Dari hasil klastering iterasi 3 menunjukan bahwa nilai *centroid* iterasi 4 sama dengan *centroid* iterasi 3 sehingga hasil klastering iterasi 4 hasil nya akan sama dengan hasil klastering iterasi 3.

Selain itu perlu membandingkan hasil perhitungan K-Means dengan metode statistik deskriptif sederhana untuk melakukan perbandingan. Perbandingan manual berdasarkan penjumlahan nilai variabel X dan Y. Wilayah dengan nilai skor tertinggi dikategorikan sebagai daerah sangat rawan, nilai sedang sebagai daerah rawan, dan nilai terendah sebagai daerah tidak rawan.

Berdasarkan data pada Tabel 15 menunjukan bahwa hasil klastering pada metode sederhana berbeda dengan hasil klastering menggunakan algoritma K-Means. Hal ini dikarenakan metode K-Means mempertimbangkan pola jarak multivariat, bukan hanya nilai total, sehingga mampu menangkap struktur data yang tidak terlihat oleh metode ranking sederhana.

Tabel 15 Perbandingan k-means dengan metode ranking sederhana

No	Wilayah	X	Y	Metode Sederhana		Metode Algoritma K-Means iterasi			Cluster	Potensi
				Jumlah	Potensi	Jarak ke C1	Jarak ke C2	Jarak ke C3		
1	010. Sukolilo	0.25	0.3	0.55	Sangat Rawan	0.3476	0.1271	0.1122	C3	Tidak Rawan
2	020. Kayen	0.333	0.2667	0.5997	Sangat Rawan	0.2791	0.0528	0.2009	C2	Rawan
3	030. Tambakromo	0.2889	0.6	0.8889	Sangat Rawan	0.6132	0.3882	0.3065	C3	Tidak Rawan
4	040. Winong	0.2	0.4	0.6	Sangat Rawan	0.4543	0.2371	0.0894	C3	Tidak Rawan
5	050. Pucakwangi	0.2	0.2	0.4	Tidak Rawan	0.2941	0.1511	0.1429	C3	Tidak Rawan
6	060. Jaken	0.3	0.2	0.5	Rawan	0.2308	0.0526	0.2046	C2	Rawan
7	070. Batangan	0.3778	0	0.3778	Tidak Rawan	0.0378	0.2186	0.4057	C1	Sangat Rawan
8	080. Juwana	0.4533	0	0.4533	Rawan	0.0378	0.2402	0.4539	C1	Sangat Rawan
9	090. Jakenan	0.4333	0	0.4333	Rawan	0.0178	0.2276	0.4401	C1	Sangat Rawan
10	100. Pati	0.3	0.2	0.5	Sangat Rawan	0.2315	0.0531	0.2066	C2	Rawan
11	110. Gabus	0.3647	0	0.3647	Tidak Rawan	0.0508	0.2174	0.3882	C1	Sangat Rawan
12	120. Margorejo	0.4667	0.2	0.6667	Sangat Rawan	0.2068	0.1177	0.3512	C2	Rawan
13	130. Gembong	0.2	0.2667	0.4667	Rawan	0.3271	0.1594	0.0863	C3	Tidak Rawan
14	140. Tlogowungu	0	0.4	0.4	Rawan	0.5678	0.4029	0.1584	C3	Tidak Rawan
15	150. Wedarijaksa	0.6	0	0.6	Sangat Rawan	0.1845	0.3306	0.5543	C1	Sangat Rawan
16	160. Trangkil	0.4	0	0.4	Rawan	0.0155	0.2217	0.4103	C1	Sangat Rawan
17	170. Margoyoso	0.4	0	0.4	Rawan	0.0155	0.2217	0.4103	C1	Sangat Rawan
18	180. Gunungwungkal	0	0.28	0.28	Tidak Rawan	0.4924	0.3567	0.1504	C3	Tidak Rawan
19	190. Cluwak	0	0.2	0.2	Tidak Rawan	0.4612	0.3512	0.1892	C3	Tidak Rawan
20	200. Tayu	0.36	0	0.36	Tidak Rawan	0.0555	0.2169	0.3884	C1	Sangat Rawan
21	210. Dukuhseti	0.35	0	0.35	Tidak Rawan	0.0655	0.2167	0.3881	C1	Sangat Rawan

Hasil proses iterasi akhir berdasarkan data yang bencana banjir dan tanah longsor yang terjadi di Kabupaten Pati dengan rentang tahun dari 2018 sampai dengan 2022 menunjukan bahwa kabupaten Pati memiliki 42.86% wilayah yang berpotensi sangat rawan bencana, 19.04% wilayah yang berpotensi rawan bencana dan 38.1% wilayah yang tidak berpotensi rawan bencana.

Hasil pengelompokan wilayah pada penelitian ini merupakan zonasi berbasis riwayat kejadian banjir, bukan zonasi risiko fisik banjir. Klasifikasi wilayah dilakukan berdasarkan data historis kejadian banjir yang direpresentasikan oleh variabel X dan Y, sehingga kategori tidak rawan, rawan, dan sangat rawan mencerminkan frekuensi serta pola kejadian banjir yang pernah terjadi pada masing-masing wilayah. Dengan demikian, hasil zonasi ini menunjukkan tingkat kerentanan wilayah berdasarkan pengalaman kejadian masa lalu, bukan berdasarkan analisis fisik seperti topografi, curah hujan, atau kapasitas aliran sungai.

Hasil yang diperoleh dari *clustering* dapat divisualisasikan dalam bentuk peta menggunakan WebGis (Qgis). Pada aplikasi Qgis tersebut tabel atribut kolom potensi digunakan sebagai acuan untuk pola pewarnaan dalam peta. Sebelum *export* ke web *popup fields* dipilih dan digunakan untuk menampilkan informasi label wilayah saat kursor men-hover suatu wilayah peta seperti yang ditampilkan pada Gambar 3. Warna hijau mewakili daerah dengan kategori tidak rawan bencana. Kemudian warna kuning menunjukan daerah dengan kategori rawan bencana. Sedangkan warna oranye menunjukan daerah dengan kategori sangat rawan bencana. Saat kursor diarahkan kesalahsatu wilayah maka akan menampilkan informasi nama kecamatan, label potensi bencana dari kecamatan tersebut. Dengan adanya visualisasi pemetaan, hasil analisis dapat terlihat dengan jelas wilayah mana saja yang memiliki potensi bencana yang lebih tinggi.



Gambar 3 Peta rawan bencana

5 Kesimpulan

Dari hasil penelitian, algoritma K-means dapat di terapkan untuk menentukan wilayah yang berpotensi bencana alam banjir dan tanah longsor menjadi 3 *Cluster*: tidak rawan (C3), rawan (C2) dan sangat rawan (C1). *Cluster* C1 dicirikan oleh banyaknya wilayah desa yang terkena banjir sehingga menjadikan wilayah tersebut daerah sangat rawan bencana banjir. Hasil klastering berdasarkan data bencana banjir dan tanah longsor pada periode 2018 - 2022, terdapat 9 wilayah (42.86%) di Kabupaten Pati paling sering mengalami bencana yaitu Kecamatan Dukuhseti, Kecamatan Tayu, Kecamatan Gabus, Kecamatan Batangan, Kecamatan Trangkil, Kecamatan Margoyoso, Kecamatan Jakenan, Kecamatan Juwana, Kecamatan Wedarijaksa. Beberapa kecamatan masuk dalam *cluster* rawan bencana (19.04%) adalah Kecamatan Jaken, Kecamatan Pati, Kecamatan Kayen, Kecamatan Margorejo. Beberapa wilayah yang masuk dalam *cluster* tidak rawan bencana (38.1%) meliputi Kecamatan Cluwak, Kecamatan Gunungwungkal, Kecamatan Tlogowungu, Kecamatan Pucakwangi, Kecamatan Gembong, Kecamatan Winong, Kecamatan Sukolilo. Berdasarkan hasil klastering tersebut, disarankan pada penelitian selanjutnya untuk menambahkan beberapa variabel yang lebih kompleks seperti jumlah korban, curah hujan, jumlah sungai dan jenis bencana lainnya.

Referensi

- [1] A. C. Syahwanes, E. A. Saepudin, R. Utari, A. Hayatunnisa, I. K. Ain, and M. R. Sandy, "Analisis Kebijakan Manajemen Risiko Bencana di Daerah Rawan Bencana," *J. Multidiscip. Inq. SCI. Technol. Educ. Res.*, Vol. 2, No. 1, pp. 1468–1478, 2025.
- [2] W. F. Adiguna and H. Wahyono, "Interaksi Kelembagaan dalam Pengelolaan dan Pemanfaatan Kawasan Ekologis Bentang Alam Karst Sukolilo di Kabupaten Pati," *J. Biol. Educ.*, Vol. 4, No. 1, p. 1, Jun. 2021, DOI: 10.21043/job.v4i1.10249.
- [3] H. T. A. Khoirun Nisa, "Strategi Mitigasi Bencana oleh Badan Penanggulangan Bencana Daerah (BPBD) dalam Upaya menanggulangi Banjir di Kabupaten Pati," *J. Polit. Gov. Stud.*, Vol. 32, No. 3, pp. 167–186, 2021.
- [4] M. R. Andhes Ta and S. Raha Yu, "Kajian Risiko Banjir di Kabupaten Pati berbasis Sistem Informasi Geografis," *Perenc. Wil. Dan Kota*, Vol. 6, No. 3, pp. 202–212, 2017, [Online]. Available: <http://ejournal3.undip.ac.id/index.php/pwk>
- [5] M. Solahuddin, "SIG untuk Memetakan Daerah Banjir dengan Metode Skoring dan Pembobotan (Studi Kasus Kabupaten Jepara)," 2014.
- [6] D. Desmawan, A. Nuraisah, A. Mumtaz, M. F. Fadhilah, F. S. Aufa, and L. Malihati, "Dampak Pengalihan Fungsi Lahan Pertanian menjadi Lahan Permukiman dan Industri di Kawasan Kabupaten Bekasi," *Bursa J. Ekon. dan Bisnis*, Vol. 3, No. 3, pp. 115–121, Dec. 2024, DOI: 10.59086/jeb.v3i3.572.
- [7] A. Asrul, "Pengaruh Alih Fungsi Lahan terhadap Kejadian Banjir di Kelurahan Leato Selatan Kecamatan Kota Timur Kota Gorontalo," *Geodika J. Kaji. Ilmu dan Pendidik. Geogr.*, Vol. 9, No. 1, pp. 52–61, Jan. 2025, DOI: 10.29408/geodika.v9i1.29256.
- [8] M. Berliandaldo, A. Prasetyo, and V. P. I. Sakti, "Kebijakan Pengembangan Pariwisata Tangguh Bencana melalui Kolaborasi dan Manajemen Pariwisata Kebencanaan Terintegrasi," *J. Abdimas Pariwisata*, Vol. 4, No. 1, pp. 62–77, Jan. 2023, DOI: 10.36276/jap.v4i1.422.
- [9] B. D. Y. Fina Faizana, Arief Laila Nugraha, "Pemetaan Risiko Bencana Tanah Longsor Kota Semarang," *J. Geod. Undip*, Vol. 4, p. 35, 2004.
- [10] N. Hendrastuty, "Penerapan Data Mining menggunakan Algoritma K-Means Clustering dalam Evaluasi Hasil Pembelajaran Siswa," *J. Ilm. Inform. dan Ilmu Komput.*, Vol. 3, No. 1, pp. 46–56, Mar. 2024, DOI: 10.58602/jima-ilkom.v3i1.26.
- [11] I. Hidayat, E. Darnila, and Y. Afrillia, "Clustering Zonasi Daerah Rawan Bencana Alam di Kabupaten Mandailing Natal menggunakan Algoritma K-Means," *G-Tech J. Teknol. Terap.*, Vol. 7, No. 3, pp. 1218–1226, Jul. 2023, DOI: 10.33379/gtech.v7i3.2880.
- [12] M. F. A. Faqih and E. Mailoa, "Clustering Zonasi Daerah Rawan Bencana Alam Provinsi Jawa Tengah menggunakan Algoritma K-Means dan Library Geopandas," *IT-Explore J. Penerapan Teknol. Inf. dan Komun.*, Vol. 4, No. 1, pp. 116–127, Feb. 2025, DOI: 10.24246/itexplore.v4i1.2025.pp116-127.

- [13] M. Murdiaty, A. Angela, and C. Sylvia, "Pengelompokan Data Bencana Alam berdasarkan Wilayah, Waktu, Jumlah Korban dan Kerusakan Fasilitas dengan Algoritma *K-Means*," *J. Media Inform. Budidarma*, Vol. 4, No. 3, p. 744, Jul. 2020, DOI: 10.30865/mib.v4i3.2213.
- [14] W. Ananda, I. Hartami Santi, and S. Kirom, "Penerapan Algoritma *K-Means Clustering* dalam Pengelompokan Arsip SKCK," *JATI (Jurnal Mhs. Tek. Inform.*, Vol. 6, No. 2, pp. 861–867, Nov. 2022, DOI: 10.36040/jati.v6i2.5762.
- [15] L. M. Harahap, W. Fuadi, L. Rosnita, E. Darnila, and R. Meiyanti, "Klastering Sayuran Unggulan menggunakan Algoritma *K-Means*," *J. Tek. Inform. dan Sist. Inf.*, Vol. 8, No. 3, Dec. 2022, DOI: 10.28932/jutisi.v8i3.5277.
- [16] R. Risawandi and Y. Afrillia, "Geographic Information System Mapping of Criminality Villed Areas in Lhokseumawe using *K-Means Method*," *J. INFORMATICS Telecommun. Eng.*, Vol. 5, No. 2, pp. 442–451, Jan. 2022, DOI: 10.31289/jite.v5i2.6265.
- [17] Z.-R. Peng and M.-H. Tsou, *Internet GIS: Distributed Geographic Information Service for the Internet and Wireless Networks*. John Wiley & Sons, 2023.
- [18] R. Hermiati, A. Asnawati, and I. Kanedi, "Pembuatan *E-Commerce* pada Raja Komputer menggunakan Bahasa Pemrograman PHP dan *Database MySQL*," *J. Media Infotama*, Vol. 17, No. 1, Feb. 2021, DOI: 10.37676/jmi.v17i1.1317.
- [19] D. Freedman, R. Pisani, and R. Purves, *Statistics*. W. W. Norton & Company, 2007.
- [20] A. Harmain, P. Paiman, H. Kurniawan, K. Kusrini, and D. Maulina, "Normalisasi Data untuk Efisiensi *K-Means* pada Pengelompokan Wilayah berpotensi Kebakaran Hutan dan Lahan berdasarkan Sebaran Titik Panas," *Tek. Teknol. Inf. dan Multimed.*, Vol. 2, No. 2, pp. 83–89, Jan. 2022, DOI: 10.46764/teknimedia.v2i2.49.