ITEJ
Indonesian Technology and Education Journal
E-ISSN: 2985-8216; P-ISSN: 2985-8208
https://journal.diginus.id/index.php/ITEJ

Flood Disaster Risk Zone Mapping in North Bolaang Mongondow Regency Using K-Means Algorithm

Budiyanto Ahaliki[1], Salahuddin Olii[2], Joudi Fahrezi Binolombangan[3*], Nikmasari Pakaya[4], Rahmat Taufik R.L Bau[5], Hermila A[6], Wildan[7]

[1,23,4,5,6] *Informatic Engineering, Faculty of Engineering, University State of Gorontalo. Indonesia*
[7] *Department of Electrical and Computer Engineering, Gorontalo State University, Gorontalo, Indonesia*
Corresponding e-mail : budiyanto@ung.ac.id

ARTICLE INFO

Keywords:

Flood Zoning Mapping
K-Means Clustering
Geographic Information System

Article History

Received: Month XX, 20xx
Revised : Month XX. 20xx
Accepted : Month XX, 20xx
(Cambria 9)

ABSTRACT

This research aims to identify and map flood-prone zones in North Bolaang Mongondow Regency using the K-Means algorithm, as well as to analyze the key factors influencing flood vulnerability. This study employs a quantitative approach with spatial analysis utilizing geospatial datasets including rainfall intensity, slope gradient, and historical flood occurrences. The data were processed using Geographic Information Systems (GIS) and analyzed using the K-Means algorithm to classify areas into three vulnerability categories: highly vulnerable (Tinggi), moderately vulnerable (Sedang), and not vulnerable (Rendah). Results indicate that the K-Means algorithm effectively clusters regions with a Silhouette Coefficient of 0.607 for k=3, outperforming k=2 (0.485). The algorithm converged at iteration 5, classifying 107 villages into: 19 (17.76%) as highly flood-prone, 80 (74.77%) as moderately flood-prone, and 8 (7.48%) as low flood-prone. A web-based GIS system was developed to visualize and disseminate the zonation results for community awareness and disaster mitigation planning.

To cite this article : Author. (20xx). Title. Information Technology Education Journal, X(X), XX-XX. Doi. Xxxx

INTRODUCTION

Indonesia is one of the most disaster-prone countries in the world, particularly vulnerable to hydrometeorological disasters such as flooding. The country's topography, dominated by mountainous and archipelagic terrain, significantly contributes to the high frequency of natural disasters across various regions. Based on data from the National Disaster Management Authority (BNPB), in February 2023 alone, 295 disaster events were recorded, with floods accounting for 125 incidents, demonstrating the persistent threat of flooding to Indonesian communities (Sulistya, 2022).

North Bolaang Mongondow Regency (Kabupaten Bolaang Mongondow Utara) in North Sulawesi Province is geographically positioned at coordinates 124°19'12"-124°45'36' East Longitude and 0°25'12"-0°53'24' North Latitude. The regency possesses complex geographic characteristics with varied topography ranging from lowlands to hills, several major rivers, high annual rainfall, and a recurring history of flood events, particularly during the rainy season. These conditions underscore the urgent need for systematic flood disaster zone mapping to support mitigation planning and disaster risk management (Babay et al., 2024).

Clustering is a technique for grouping data based on similarity characteristics. K-Means Clustering is among the most popular algorithms for grouping problems due to its simplicity and high accuracy in handling various data sizes. K-Means is a non-hierarchical clustering method with a simple and fast data grouping technique (Alfian, 2023). The application of K-Means to geospatial data enables effective classification of regions based on flood vulnerability parameters.

This study aims to develop a web-based Geographic Information System (GIS) that applies K-Means Clustering to map flood-prone zones in North Bolaang Mongondow Regency. The system classifies 107 villages across six sub-districts into three vulnerability categories. High (Sangat Rawan), Moderate (Rawan), and Low (Tidak Rawan), based on three primary parameters: rainfall intensity, land slope gradient, and historical flood occurrences.

METHOD

This study employs a quantitative research approach with spatial analysis, defined as a research method grounded in geospatial and statistical data analysis for examining spatial distribution patterns of phenomena (Sumaryanto, 2021). The research uses K-Means Clustering to identify and map flood-prone zones based on multiple physical and hydrological parameters across North Bolaang Mongondow Regency.

Research Location and Data

The research was conducted in North Bolaang Mongondow Regency, North Sulawesi Province, encompassing all six sub-districts: Bintauna, Bolangitan Barat, Bolangitan Timur, Kaidipang, Pinogaluman, and Sangkub. Data collection was performed in May 2025 using secondary data from the Central Bureau of Statistics (BPS), the Meteorology, Climatology, and Geophysics Agency (BMKG), and the Regional Disaster Management Agency (BPBD).

Table 1. Research Dataset (North Bolaang Mongondow Regency)

No.	Sub-district	Rainfall (mm)	Slope (°)	Flood History
1.	Bintauna	2,593.56	1.69	3.31
2.	Bolangitan Barat	2,834.10	3.72	3.00
3.	Bolangitan Timur	2,552.75	3.95	2.25
4.	Kaidipang	2,730.80	9.47	2.87
5.	Pinogaluman	2,690.95	5.41	2.41
6.	Sangkub	2,827.94	3.81	3.19

Research Design

The research design encompasses five sequential stages: (1) Spatial Data Input including slope, rainfall, and flood history data; (2) Data Preprocessing including normalization, grid sampling, and value extraction; (3) K-Means Algorithm execution including centroid initialization, cluster assignment, centroid update, and iteration until convergence; (4) Model Evaluation using Silhouette Score, WCSS Analysis, and Elbow Method; and (5) Result Validation using Ground Truth, Historical Data, and Expert Judgment.

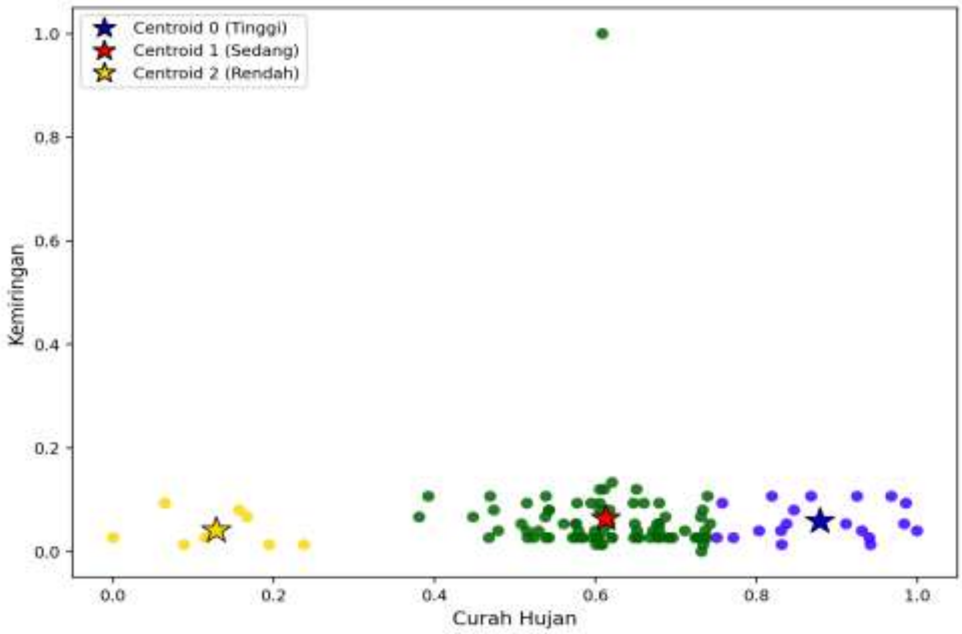

Figure 1. Research Design of Flood Zone Mapping System

Data Preprocessing

Data preprocessing consists of three steps. First, Data Integration combines data from three different sources: rainfall from BMKG, slope data from digital maps, and flood occurrence data from BPBD historical records. Second, Attribute Selection identifies relevant attributes for K-Means clustering, utilizing rainfall (mm), slope (degrees), and number of flood occurrences per year, while excluding administrative codes and district names. Third, Data Normalization applies Standard Scaling to transform all variables to mean=0 and standard deviation=1, ensuring each variable contributes equally to the clustering process.

K-Means Algorithm Implementation

The K-Means algorithm was implemented with k=3 clusters representing three flood vulnerability zones: Zone 1 (Tinggi/High), Zone 2 (Sedang/Moderate), and Zone 3 (Rendah/Low). Initial centroids were selected based on maximum, median, and minimum dataset values. Euclidean Distance was used as the distance metric:

$$d(x,y) = \sqrt{\Sigma_i}(x_i - y_i)^2$$

Initial centroids were established as follows: Centroid 1 (High Zone): C1=[884, 10, 35]; Centroid 2 (Moderate Zone): C2=[707, 13, 4]; Centroid 3 (Low Zone): C3=[674, 13, 6]. Iterations were performed until convergence, evaluated using Silhouette Coefficient analysis for both k=2 and k=3 to determine the optimal cluster count.

System Development

A web-based GIS system was developed using Python (Flask framework), MySQL database, and Leaflet.js for map visualization. The system architecture was designed to accommodate two main user roles: Administrator (BPBD staff) for data management and cluster processing, and Public Users (community members) for accessing flood risk information. The system was tested using Blackbox and Whitebox testing methodologies.

E-ISSN: 2985-8216
P-ISSN: 2985-8208

RESULTS AND DISCUSSION

K-Means Clustering Results

The K-Means algorithm was applied to the dataset of 107 villages across six sub-districts. The algorithm underwent seven iterations, converging definitively at iteration 5 where no cluster membership changes occurred. Table 2 summarizes the Euclidean Distance results at convergence (Iterations 5, 6, and 7 produced identical results):

Table 2. K-Means Clustering Results – Final Iteration

No.	Sub-district	K1 (High)	K2 (Moderate)	K3 (Low)	Cluster
1	Bintauna	443.08	187.26	550.07	Moderate
2	Bolangitan Barat	224.51	152.97	784.11	Moderate
3	Bolangitan Timur	483.89	214.19	512.77	Moderate
4	Kaidipang	321.84	103.65	681.08	Moderate
5	Pinogaluman	351.93	114.55	640.96	Moderate
6	Sangkub	271.70	287.34	798.94	High

At convergence, five sub-districts (Bintauna, Bolangitan Barat, Bolangitan Timur, Kaidipang, and Pinogaluman) are classified as Moderate flood risk (K2 distance consistently lowest), while Sangkub is classified as High flood risk. Notably, Kaidipang records the lowest K2 average distance (103.65), confirming its strong centroid proximity to the moderate zone. Sangkub's transition to High risk emerged at Iteration 2 when its K1 value (282.71) became lower than K2 (284.45), indicating higher rainfall and flood history characteristics.

Overall, the K-Means algorithm successfully classified all 107 villages into three flood vulnerability zones: 19 villages (17.76%) as High flood-prone (Sangat Rawan), 80 villages (74.77%) as Moderate flood-prone (Rawan), and 8 villages (7.48%) as Low flood-prone (Tidak Rawan).

Cluster Visualization

The clustering results were visualized in both 2D and 3D scatter plots to illustrate the spatial distribution of flood vulnerability zones across North Bolaang Mongondow Regency.

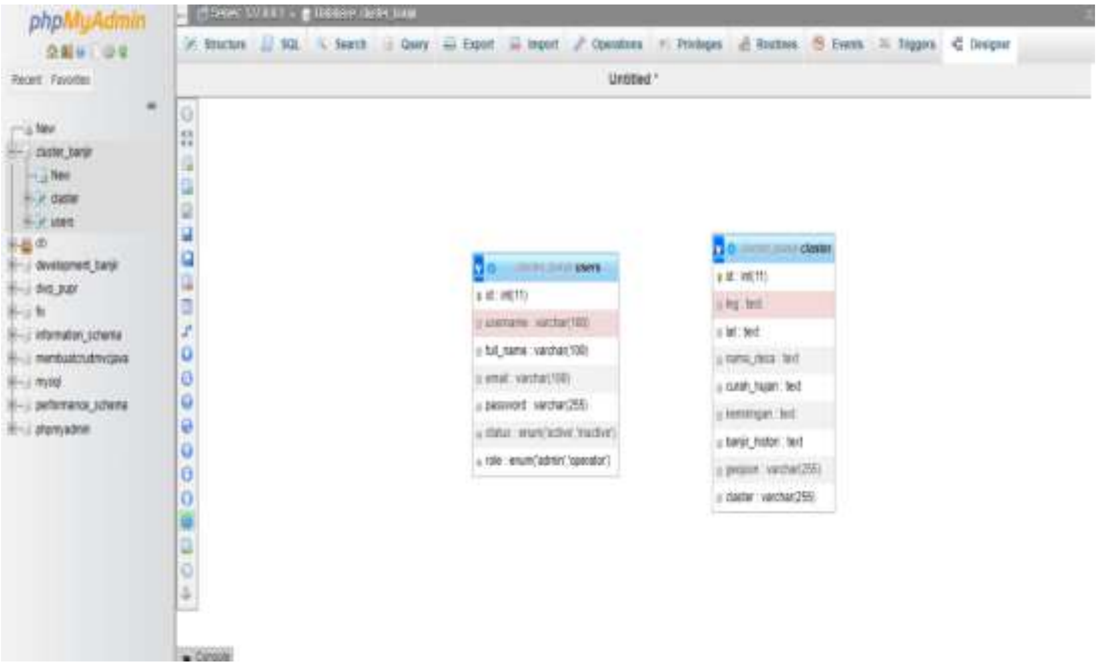

Figure 2. 2D Visualization of K-Means Clustering Results – Flood Vulnerability Zones of North Bolaang Mongondow Regency

Based on the 2D visualization (Figure 2), data points normalized to the 0–1 range show clear cluster separation. Star markers (★) indicate the final centroid positions for each cluster: Centroid 0 (blue) for High Zone, Centroid 1 (red) for Moderate Zone, and Centroid 2 (yellow) for Low Zone. The Moderate Zone (green) dominates the mid-to-high rainfall range, while the Low Zone (purple) concentrates at high rainfall but with low slope gradient.

Figure 3. 3D Visualization of K-Means Clustering Results – Flood Vulnerability Zones

The 3D visualization (Figure 3) provides a clearer perspective of the three-dimensional data distribution. The High Zone (pink) is positioned at high rainfall (>2,900 mm). The Moderate Zone (dark blue) dominates with the most data points, spanning rainfall of 2,400–2,900 mm. The Low Zone (yellow) occupies lower rainfall ranges (<2,300 mm). The three-dimensional representation confirms clear cluster separation along all three parameters simultaneously.

Silhouette Coefficient Evaluation

To determine the optimal number of clusters, Silhouette Coefficient analysis was performed for both k=2 and k=3:

Table 3. Silhouette Coefficient Comparison

k Value	Cluster Configuration	Total Data	Total S(i)	Average Silhouette
k = 2	High | Moderate+Low	107	51.92183	0.48525
k = 3	High | Moderate | Low	107	64.97456	0.60724

The Silhouette Coefficient for k=3 (0.607) significantly outperforms k=2 (0.485). Silhouette values closer to 1 indicate better-defined clusters with strong intra-cluster cohesion and clear inter-cluster separation. The result confirms that classifying flood-prone areas into three zones (High, Moderate, Low) produces more representative and statistically valid groupings than two-zone classification, providing greater granularity for disaster mitigation planning.

Web-Based GIS System

A web-based Geographic Information System was developed to implement and visualize the K-Means clustering results. The system architecture is based on the Data Flow Diagram (DFD) framework encompassing data collection, processing, and output presentation stages.

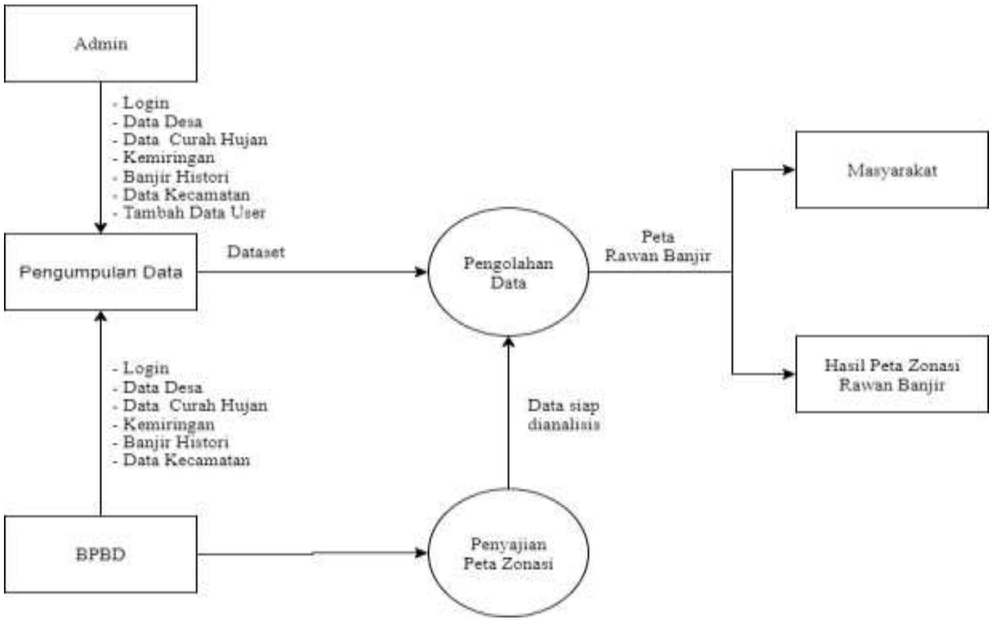

Figure 4. Context Diagram of Flood Zone Mapping GIS System

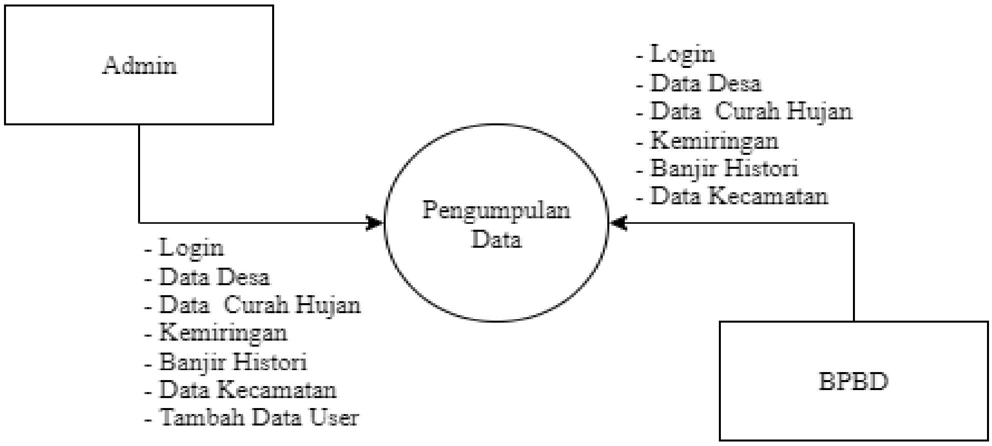

Figure 5. Data Flow Diagram (DFD) Level 0

The system implements Python (Flask) as the backend framework with MySQL for data management. The K-Means implementation utilizes scikit-learn's KMeans with MinMaxScaler normalization. Key system features include:

1. Admin Module: Full CRUD operations for spatial data (longitude, latitude, rainfall, slope, flood history per village), cluster processing execution, and result validation.

2. Public Module: Interactive flood risk map with sub-district filtering, flood zone statistics dashboard, and risk analysis table.

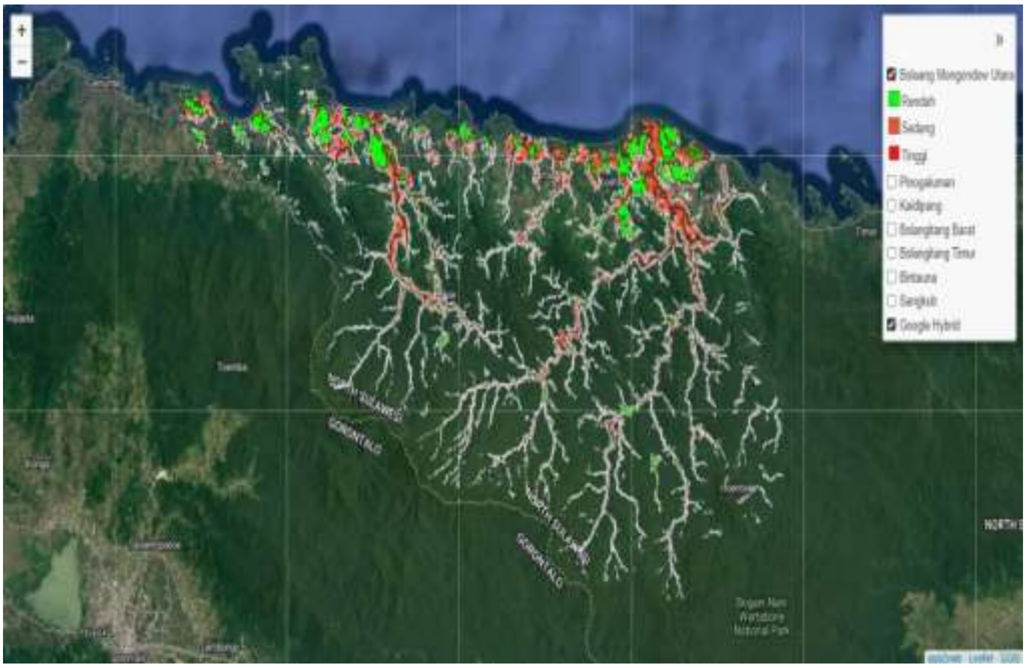

Figure 6. Flood Risk Zone Map Display North Bolaang Mongondow Regency

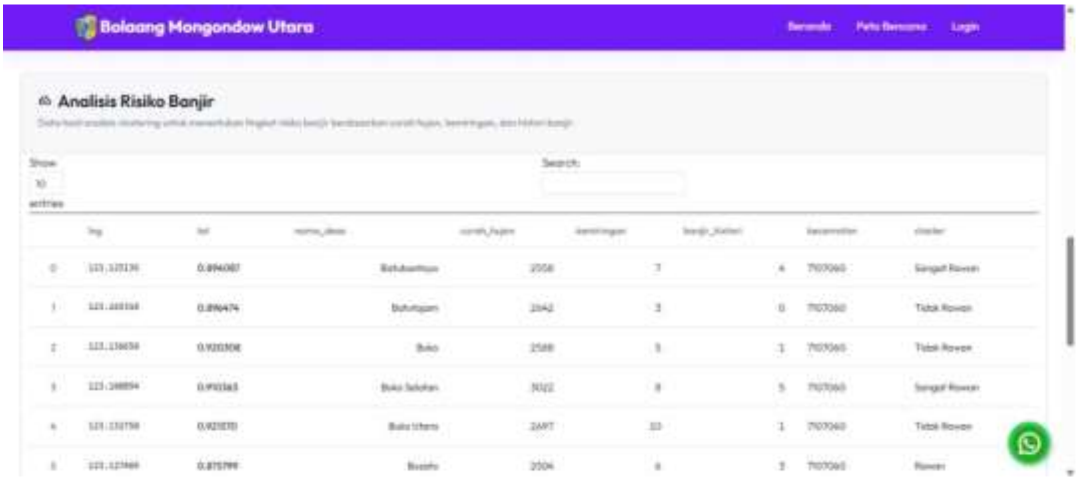

Figure 7. Flood Risk Analysis Display per Village

The flood zone map (Figure 6) displays three vulnerability levels using color coding: green for Low risk, orange for Moderate risk, and red for High risk, with sub-district filtering capability. The risk analysis table (Figure 7) presents detailed information per village including coordinates, rainfall, slope, flood history, and risk category, enabling granular decision-making for disaster mitigation.

System Testing

The system was validated through Blackbox testing (functional testing of 10 test cases including login, data input, synchronization, cluster processing, and user management—all returning successful results) and Whitebox testing (flowchart and flowgraph analysis yielding Cyclomatic Complexity V(G) = 12 - 11 + 2 = 3, with three independent paths identified).

Table 4. Blackbox Testing Results Summary

Test Case	Expected Result	Status
Open Application	Display main page without error	Success
Admin Login	Redirect to dashboard	Success
Empty Login Fields	Show error message	Success
Dashboard Search	Display village and risk info	Success
Input Data	Show "Data Saved" message	Success
Data Synchronization	Sync successful	Success
Cluster Processing	Process completed	Success
Reset Data	Data reset successful	Success
Add User	User added successfully	Success
Delete User	User deleted successfully	Success

All 10 functional test cases passed successfully, confirming system reliability. The Whitebox analysis identified three independent execution paths, verifying complete logical coverage of the K-Means processing code.

CONCLUSIONS

This research successfully developed a web-based Geographic Information System for flood disaster zone mapping in North Bolaang Mongondow Regency using K-Means Clustering algorithm. The K-Means algorithm classified 107 villages into three flood vulnerability categories with a Silhouette Coefficient of 0.607, converging at iteration 5. Results show that 19 villages (17.76%) are classified as High flood-prone, 80 villages (74.77%) as Moderate flood-prone, and 8 villages (7.48%) as Low flood-prone. Sangkub sub-district exhibited the highest flood vulnerability characteristics, while Kaidipang showed the strongest affinity to the Moderate zone centroid.

The developed GIS system successfully integrates K-Means results into an accessible web platform supporting both administrative management (BPBD) and public information access. System testing confirmed full functional reliability through Blackbox and Whitebox methodologies. Future development should incorporate real-time rainfall monitoring integration, mobile accessibility, and expanded datasets to enhance zonation accuracy and disaster early warning capabilities.

AKNOWLEDGMENT (IF ANY)

The author expresses gratitude to the supervisors Salahudin Olii, S.T., M.T. and Budiyanto Ahaliki, S.Si., M.Kom for their guidance, to BMKG, BPS, and BPBD of North Bolaang Mongondow Regency for providing research data, and to the Faculty of Engineering, Universitas Negeri Gorontalo for institutional support.

REFERENCES

Alfian, I. (2023). Penerapan Metode K-Means dalam Melakukan Pengelompokan Bencana Alam di Indonesia Dilakukan dengan Memanfaatkan Teknik Text Mining. Jurnal Algoritma, 20(1), 139–147.

Babay, S. S., Tungka, A. E., & Moniaga, I. L. (2024). Partisipasi Masyarakat dalam Mitigasi Bencana Banjir di Kabupaten Bolaang Mongondow Utara. Jurnal Fraktal, 6(2), 9–16.

Haviluddin, H., et al. (2021). Implementasi Metode K-Means Untuk Pengelompokkan Rekomendasi Tugas Akhir. Informatika Mulawarman: Jurnal Ilmiah Ilmu Komputer, 16(1). https://doi.org/10.30872/jim.v16i1.5182

Hidayat, I., Darnila, E., & Afrillia, Y. (2023). Clustering Zonasi Daerah Rawan Bencana Alam di Kabupaten Mandaling Natal Menggunakan Algoritma K-Means. Jurnal Teknologi Terapan, 7(32), 1218–1226.

Norshahlan, M., Jaya, H., & Kustini, R. (2023). Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokan Data Calon Siswa Baru. Jurnal Sistem Informasi TGD, 2(6), 1042–1053.

Niode, D. F., Rindengan, Y. D. Y., & Karouw, S. D. S. (2021). Geographical Information System (GIS) untuk Mitigasi Bencana Alam Banjir di Kota Manado. E-Journal Teknik Elektro dan Komputer, 5(2), 14–20.

Shalih, O., W. Adi, A., & Wiguna, S. (2023). Risiko Bencana Indonesia: Memahami Risiko Sistemik di Indonesia. Pusat Data, Informasi dan Komunikasi Kebencanaan BNPB.

Sugiyono, B. (2022). Metode Penelitian Pendidikan (Pendekatan Kuantitatif, Kualitatif, dan R & D). Alfabeta.

Sulistya, W. (2022). Belajar dari Kejadian Bencana Alam Sepanjang Tahun 2021. Jurnal Widya Climago, 4(2), 84–90.

Sumaryanto. (2021). Metode Penelitian Bidang Komputer. Yayasan Prima Agus Teknik.

Wibowo, Y. A., Dewi, R. P., & Ronggowulan, L. (2020). Penguatan Literasi Mitigasi Bencana Angin Puting Beliung. Jurnal Warta LPM, 23(2), 168–179.