JURNAL TEKNOLOGI REKAYASA INFORMASI DAN KOMPUTER
E-ISSN: 2797-1724, DOI: 10.30811/jtrik.v9i1.8756
Vol. 9 No. 2 2026 pp. 46-54

Design And Development of A Flood-Prone Area Mapping System Using The K-Means Clustering Method Based On Web (Case Study: Lhoksukon District)

Nurmaisuri Suana[1], M. Khadafi[2]*, Mulyadi[3]

[12,3] Department of Computer and Information Technology, Politeknik Negeri Lhokseumawe, Indonesia

*Corresponding Author: mkhadafi@pnl.ac.id

Article info: Received 10/10/2025, Revised 15/11/2025, Accepted 27/02/2026

Abstract

Natural disasters are a form of natural events that result in a major impact on human populations. The natural disaster that frequently occurs in Lhoksukon District is flooding. Flooding is caused by continuous rainfall. Based on data from BPS (Central Statistics Agency) of North Aceh, the rainfall height in 2021 averaged 152.19 mm/month. High rainfall has caused many areas to be affected by floods and experience many losses, including disrupted road access, submerged houses, a paralyzed economy, and even loss of life. Based on data from BPBD (Regional Disaster Management Agency), in 2018 there were 19 villages affected by floods out of 75 existing villages, while in 2022 there were 54 villages affected by floods out of 75 villages in Lhoksukon District. In this research, a flood-prone area mapping system was created using the K-Means Clustering method, where the K-Means Clustering method is used to cluster villages affected by floods using 5 variables: duration of water inundation, water height, watershed (DAS - Daerah Aliran Sungai), elevation, and land cover. Based on the test results that have been conducted, there are 2 villages in the green cluster, 52 villages in the yellow cluster, and 21 villages in the red cluster. The results of this clustering are digitalized into a Geographic Information System using the Mapbox API. The digital map displayed shows the area in Lhoksukon District divided by village with green zone identification for non-vulnerable level, yellow zone for vulnerable level, and red zone for highly vulnerable level.

Kata kunci: Flood Disaster, Lhoksukon District, Region Mapping, K-Means Clustering

1. Introduction

Flooding is one of the natural disasters where water flows outside water bodies such as rivers. Flooding is considered a disaster because this water flow directly impacts life, such as environmental damage, property loss, psychological impacts, and even loss of life [1]. Flood disasters have become a global issue where almost every year, especially during the rainy season, several regions will experience flood disasters. This is due to changing environmental conditions, either rapidly or slowly, caused by various factors, one of which is human activities that conflict with normal environmental conditions [2]. Lhoksukon District is an area that is classified as frequently experiencing flood disasters and is considered a highly vulnerable area in North Aceh Regency. According to reports from BPBD (Regional Disaster Management Agency) of North Aceh, in 2018 several villages in Lhoksukon District were hit by floods affecting 19 villages out of 75 existing villages, while in 2020 Lhoksukon District was hit by floods affecting 54 villages out of 75 villages in Lhoksukon District [3]. Along with development, a new technology has emerged called Geographic Information System (GIS), which is an information system used to manage, present, and analyze information related to geographic matters. Over time, GIS has been developed in such a way that it can be used to help communities search for desired locations or places [4].

A study applying the k-means clustering method was conducted by Muhammad Ali Hasyimi et al. with the title "Geographic Information System for Underprivileged Residents in Karangbesuki Village Using the K-Means Clustering Method." The application of the k-means clustering method for mapping underprivileged areas in Karangbesuki Village was successfully implemented into the system so that the system can group underprivileged residents [5]. Another research related to this study is research conducted by Robi Armando et al. with the title "Use of DEM SRTM Data for Mapping Flood-Prone Areas in Lhoksukon District, North Aceh Regency." In this research, only the area of flood zones was studied using 2 parameters: slope and elevation. After testing, the results obtained showed that the total area of flood-prone areas in Lhoksukon District is 12,187 hectares or 89% of the area of Lhoksukon District [3].

Therefore, mapping of flood-prone areas is greatly needed to assess the level of flood vulnerability in Lhoksukon District, as well as to identify which villages fall into the categories of non-flood-prone, flood-prone, and highly flood-prone. This flood vulnerability map is part of an early warning system for flood hazards and risks. In this research, a method is needed to group data based on existing values and characteristics. The use of the k-means clustering method is applied based on the principle of sorting analyzed data into groups using data taken from BPBD (Regional Disaster Management Agency) of North Aceh Regency. Mapping of flood-prone areas uses 3 clusters: non-flood-prone, flood-prone, and highly flood-prone, according to criteria of water height, duration of water inundation, watershed (DAS - Daerah Aliran Sungai), elevation, and land cover. The creation of this system is greatly needed in Lhoksukon District, which is frequently affected by flood disasters, to provide a new way for the community to have adequate preparedness before floods occur, so that the effects of flood disasters can be minimized and support anticipation of loss of life.

2. Methods

2.1. System Development Technique

Application development is carried out through several stages, such as conducting data requirements analysis, functional requirements analysis of the application, and application design.

2.2. Data Requirements Analysis

One way to implement data is by conducting data requirements analysis. The data requirements for the flood-prone area mapping application using the k-means clustering method are variable data used for the calculation process. The data used includes duration of water inundation data, water height data, watershed data (DAS), elevation data, and land cover data.

2.3. Functional Requirements Analysis

Functional requirements are processes that exist in every entity involved in the application. The requirements needed in designing the flood-prone area mapping system are as follows:

1. Admin Functional Requirements
   Several requirements that can be performed by the admin in this system are as follows:
   a. Log in to the system
   b. Log out of the system
   c. View, add, edit, and delete village (gampong) data
   d. View, add, edit, and add victim data
   e. View, add, edit, and add initial cluster data
   f. Perform k-means clustering calculation process
   g. View maps of village clustering results based on zones
   h. View zone summary and can export zone summary datah.

2. User Functional Requirements
   Several requirements that can be performed by users in this system are as follows:
   a. Register in the system
   b. Log in to the system
   c. Log out of the system
   d. View village (gampong) data
   e. View victim data
   f. View maps of village clustering results based on zones
   g. View summary and export zone summary data

2.4. System Flowchart Design

K-Means Clustering Method Flowchart

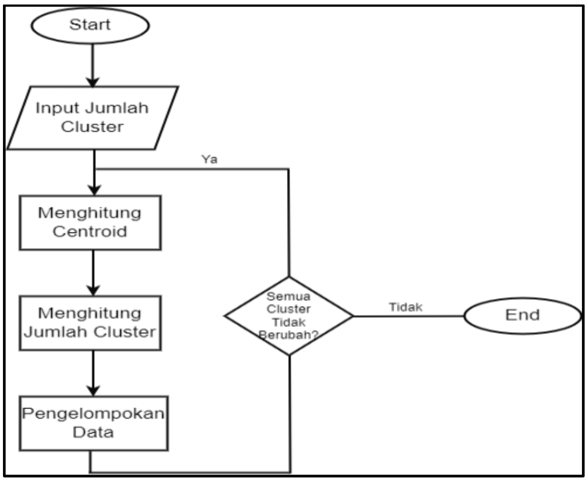


Figure 1. K-Means Clustering Method Flowchart

As shown in Figure 1, the K-Means Clustering method in performing the clustering process starts by determining the number of clusters, up to identifying whether there are any objects that move or not. If there is object movement, then the calculation restarts with new centroids, until no objects move. The following are the steps of K-Means clustering.

1. The first step is to determine k as the number of clusters.
2. Create random values for the initial centroids equal to the number of k.
3. Calculate the distance for each input data point to each respective centroid, using the Euclidean Distance formula so that the closest distance from each data point to the centroid is found. The following is equation 1 for the Euclidean Distance formula.

$$d(ai,bj)\sqrt{\sum(ai-bj)2} \quad (1)$$

Where:

ai = Criteria

bj = Centroid Cluster at $-j$

4. Classify each data point based on its proximity to the centroid (smallest distance).
5. Update the centroid values. These new centroid values are obtained from the average of the corresponding cluster using the following equation 2:

$$C_{kl=}\frac{x_{1l}+x_{2l}+\cdots+x_{2l}}{p} \quad (2)$$

Where:

C_{kl} = Centroid value at- k at variable l

p = Amount of data

6. Repeat steps 2 through 5 until there are no changes in the members of each cluster. If step 6 is fulfilled, then the cluster center values from the final iteration will be used as parameters to determine the data classification.

2.5. Use Case Diagram

In developing software, it is important to create a design to understand the process flow that occurs between the user and the system. One design that can be used is by creating a use case diagram. A use case diagram is a diagram that explains the relationship between the user and the system. The use case diagram in this system has two actors: admin and user. The explanation of this stage is illustrated in Figure 2, which describes the use case diagram where there are two actors: admin and user, as well as seven use cases: Register, Login, Mapping, Village Data, Victim Data, K-Means Clustering, and Results Summary.

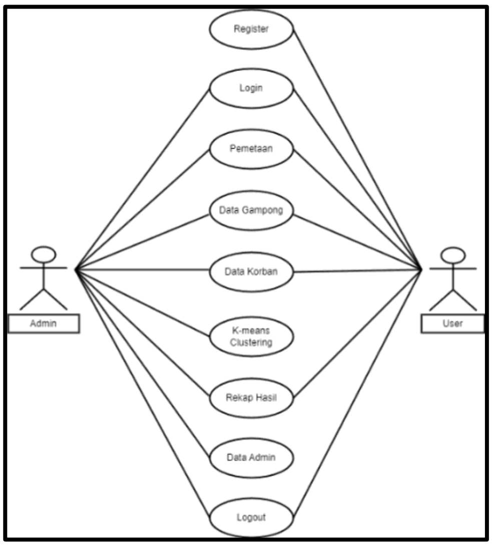


Figure 2. Use Case Diagram

Figure 2 is the system's use case diagram. From the figure above, the relationship between the user and the application is clearly visible. It can be explained that there are two actors: admin and user. The admin can perform login, view mapping, then the admin can view, delete, and add village data along with victim data. The admin can perform the K-Means Clustering calculation process and can view the zone summary and download the zone summary data. Meanwhile, the user can perform registration and login, the user can also view mapping, victim data, village data, results summary, and perform logout.

3. Result and Discussions

The results and discussion include the implementation of the K-Means Clustering method, implementation of the application user interface, and application implementation results.

3.1. Grouping Using the K-Means Clustering Method

The testing data uses 4 attributes: duration of water inundation (LGA), water height (KA), watershed area (DAS), elevation (EL), and land cover (LL). The following are the calculation steps for the K-Means Clustering method. First, determine the number of clusters. There are 3 clusters used: not flood-prone (C1), flood-prone (C2), and highly flood-prone (C3). Then initialize the data randomly based on criteria weights.

Table 1. Starting Centroid

Centroid	LGA	KA	DAS	EL	LL	Cluster
1	1	2	2	3	1	C1
2	3	3	2	2	2	C2
3	2	2	1	2	1	C3

Next, calculate the distance of each data point to the cluster center using the Euclidean distance calculation formula.

Table 2. Result of first Iteration

No	Nama Gampong	C1	C2	C3
1	Geulumpang	0	2.65	1.73
2	Mns Buloh	2.65	0	2.00
3	Alue Abee	1.73	2.00	0
4	Alue Eumpok	3.16	1.73	2.24
..	....	.....	.....	......
..	....	.....	.....	......
..	....	.....	.....	......
72	Mns Mancang	3.16	1.00	2.24
73	Leubok	3.16	1.00	2.24
74	Trieng Pantang	3.61	1.41	2.83
75	Mns Arongan Ab	2.65	2.00	1.41

From this data, the centroid centers occupying clusters C1, C2, and C3 are obtained. The data occupying each cluster will be summed, then the average value will be calculated. This value will become the new cluster center.

Table 3. New centroid at second interation

Centroid	LGA	KA	DAS	EL	LL	Cluster
1	1.67	2.00	1.33	4.67	1.00	C1
2	2.88	3.71	1.57	1.46	2.52	C2
3	1.56	2.31	1.06	1.63	1.25	C3

Next, calculate the distance of each data point from the new centroid. If the data cluster changes, then iteration needs to be performed again until the data cluster does not change.

Table 4. Result of fourth iteration

No	Nama Gampong	C1	C2	C3
1	Geulumpang	2,87	3,44	1,81
2	Mns Buloh	4,03	1,17	1,85
3	Alue Abee	3,50	2,67	0,74
4	Alue Eumpok	4,15	1,81	2,08
..	....	.....	.....	......
..	....	.....	.....	......
..	....	.....	.....	......
71	Mns Alue Drien Lb	6,26	2,59	4,13
72	Mns Mancang	4,92	1,14	1,93
73	Leubok	4,92	1,14	1,93
74	Trieng Pantang	5,22	1,05	2,48
75	Mns Arongan Ab	4,61	2,13	0,97

3.2. Maps User Interface

Figure 3 is the maps page that explains the flood-prone area map and displays the cluster results of villages that are classified as flood-prone based on the zones of each village in Lhoksukon District.

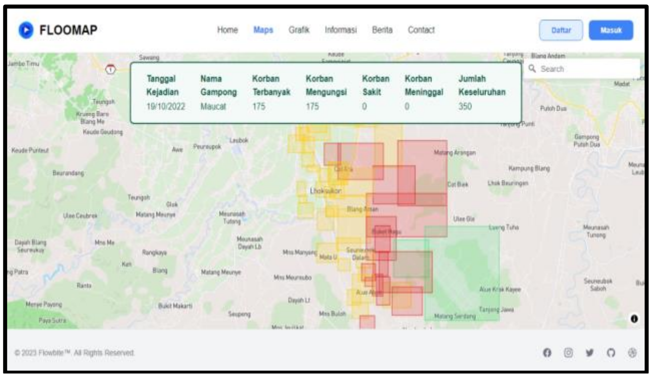


Figure 3. Maps User Interface

3.3. Chart User Interface

The graph page that displays a summary of victim data and zone cluster results can be seen in Figure 4. This graph aims to present data and information in a way that is easier to understand, can provide deeper and faster insights about the data than just numbers or text. The victim data graph is at the top, while the zone data graph is at the bottom. In the victim data graph section, each victim data that has been input will appear according to the village name and incident date from each respective village. The more victim data that is input, the more the graph data increases. Meanwhile, in the zone cluster results graph section, it only shows the total data from each cluster.

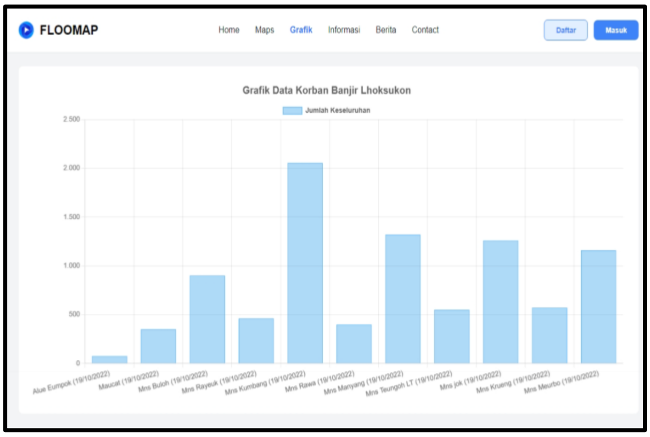


Figure 4. Chart User Interface

3.4. Information User Interface

Figure 5 is the information page that contains information sources aimed at providing knowledge and guidance related to floods to the community, such as impacts from economic, health, and education perspectives.

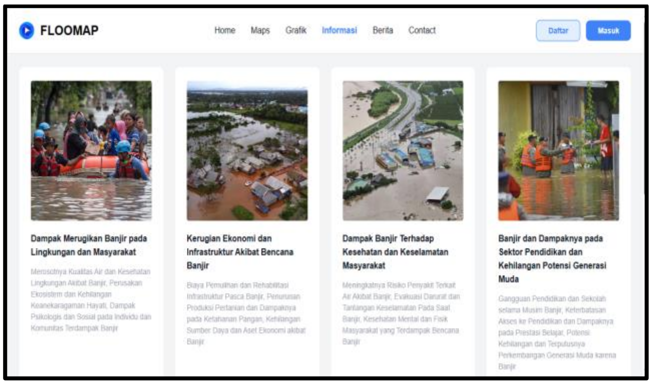


Figure 5. House 3D Object User Interface

3.5. Admin Dashboard User Interface

Figure 6 is the dashboard page used to view zone data information along with the quantity of each zone and displays a graph of the cluster results data for flood-prone areas in Lhoksukon District. The maps page displays a map of flood-prone areas in Lhoksukon District, the master data page displays village data and victim data in Lhoksukon District, while the K-Means Clustering page is the page used to perform calculations, and the last page is the zone summary which displays a summary of data from each zone.

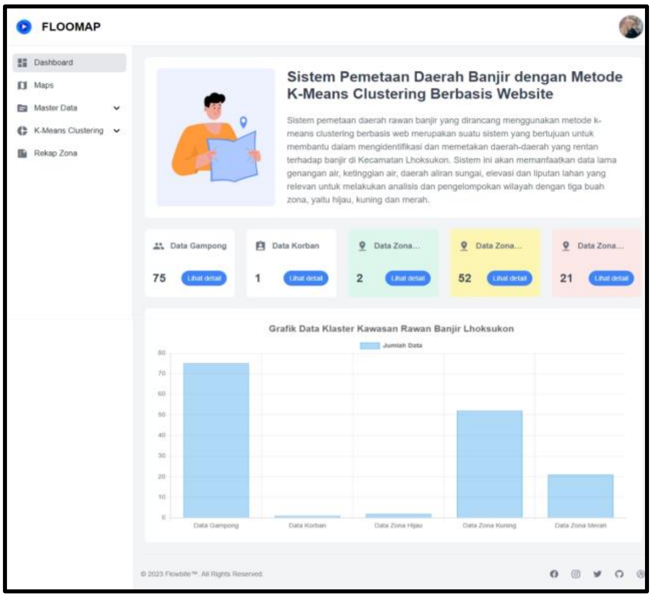
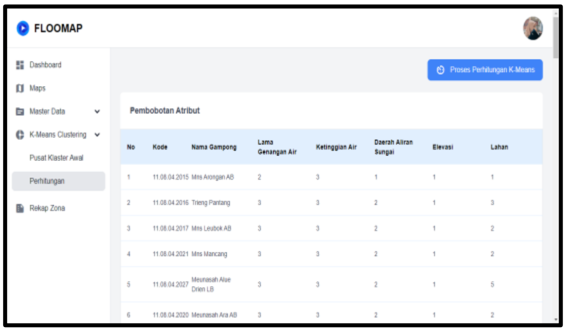


Figure 6. Admin Dashbaord User Interface

3.6. Calculation User Interface

Figure 7(a) is the initial cluster center page that contains the initial cluster data values from each variable. Figure 7(b) is the attribute value weighting process page that has been imported by the admin in the village data, which contains data from each existing variable along with the initial cluster data.

(a) Center Cluster,



(b) Weighting Process

Figure 7. Initial cluster center page

Figure 8 show the K-Means Clustering calculation process page based on the previously weighted data. The calculation process will continue running until the clusters converge. After the calculation process is complete, we can see the calculation results at which iteration they occur. In this calculation, which is finding flood-prone area zones, the calculation process using the K-Means Clustering method reached results at the 4th iteration.

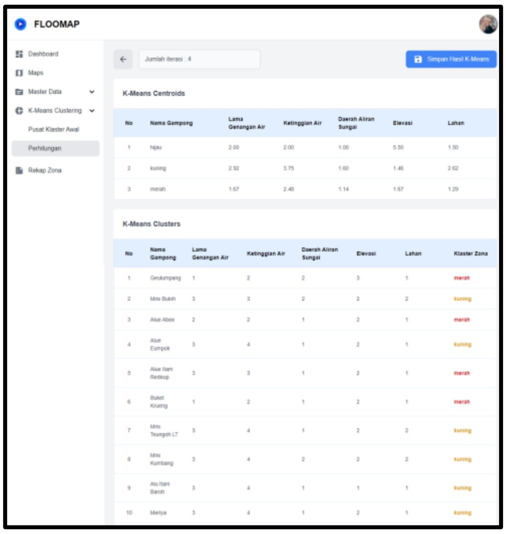


Figure 8. k-means clustering calculation

3.7. Zone User Interface

The summary page for all existing cluster zones, containing a summary of results for each existing zone along with village data and victim data from each village in Lhoksukon District is shown in Figure 9. The contents of the village data summary include village name, address, village head name (geuchik), number of houses, number of residents, number of households (KK), number of males, number of females, duration of water inundation, water height, watershed area, elevation, and land cover. Meanwhile, the contents of the victim data summary include incident date, injured victims, deceased victims, missing victims, sick victims, and total count.

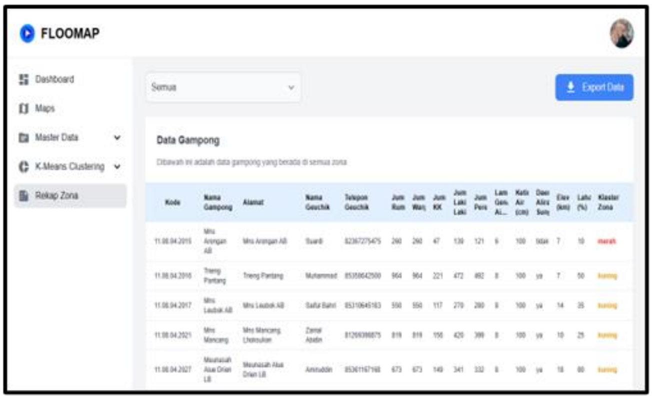


Figure 9. Summary page for all existing cluster

4. Conclussions

Based on the results and discussion that have been conducted, several conclusions can be drawn, including: This system applies the K-Means Clustering method to group flood-prone areas. The clustering results are divided into three cluster groups: green zone, yellow zone, and red zone, using five calculation criteria variables: duration of water inundation, water height, watershed area (DAS), elevation, and land cover, the results from the method calculation show that the number of clusters produced is 2 villages in the green zone, 52 villages in the yellow zone, and 21 villages in the red zone. The digital map displayed shows areas in Lhoksukon District divided by village with each zone according to the clustering results. Based on the results of black box testing using 52 test cases and white box testing, the functions and algorithms run according to the design, so it can be concluded that the system has been successfully implemented.

REFERENCE

[1] N. Zalmita, A. Fitria and A. Taher, "Tingkat kerusakan Ekonomi Pada Bencana Banjir Di Aceh Utara Tahun 2014-2019," Jurnal Geografi, vol. 19, no. 2, pp. 61-68, 2021.

[2] A. Rosyidie, "Banjir: Fakta dan Dampaknya, Serta Pengaruh dari Perubahan Guna Lahan," Jurnal Perencanaan Wilayah Dan kota, vol. 24, no. 3, pp. 241-249, 2013.

[3] R. Arnando, M. Rusdi and H. Basri, "Penggunaan Data STRM untuk Pemetaan Daerah Rawan Banjir Di Kecamatan Lhoksukon," Jurnal Ilmiah Mahasiswa Pertanian, vol. 5, no. 2, pp. 236-240, 2020.

[4] A. W. Budyastomo, "Sistem Informasi Geografis Deteksi Lokasi Kebakaran Lahan Jati Di Desa Kalijambe Kecamatan Bringin Kabupaten semarang," Interdisciplinary Journal Of Communication, vol. 1, no. 1, pp. 63-80, 2016.

[5] "Sistem Informasi Geografis Pemetaan Warga Kurang Mampu Di Kelurahan Karangbesuki Menggunakan Metode K-Means Clustering," Jurnal Mahasiswa Teknik Informatika, vol. 5, no. 1, pp. 284-290, 2021.

[6] S. H. S. Budinetro, I. S. and N. Ekarina, Pedoman Pengelolaan Bencana Banjir, Bandung: Pusat Litbang Sumber Daya air, 2014.

[7] R. Muliani, Pengenalan Tingkat Kemanisan Buah Pepaya CALLINA (Pepaya Madu) Menggunakan Pengolahan Citra Berdasarkan Warna (RGB) Dengan Metode K-Means Clustering, Lhokseumawe: Politeknik Negeri Lhkseumawe, 2016.

[8] Haviluddin, "Memahami Penggunaan UML (Unified Modeling Language)," Jurnal Informatika Mulawarman, vol. 6, no. 1, pp. 1-10, 2011.