JOURNAL OF INFORMATION TECHNOLOGY AND ITS UTILIZATION, VOLUME 7, ISSUE 1, JUNE 2024
EISSN 2654-802X ; PISSN 2985-4067
DOI: https://doi.org/10.56873/JITU.7.1.5403. SUBMITTED: DECEMBER 14, 2023; REVISED: MARCH 13, 2024; ACCEPTED: APRIL 5, 2024

Geographical Information System For Mapping Flood-Prone Areas in Manado City Using the K-Means Clustering Method

Aurelia Laurencia Koagouw[1], Debby Paseru[2*], Indah Y. Kairupan[3]

[1,2,3]Informatics Engineering Study Program, Faculty of Engineering, De La Salle Catholic University Manado
19013082@unikadelasalle.ac.id[1*], dpaseru@unikadelasalle.ac.id[2], ikairupan@unikadelasalle.ac.id[3]
*Corresponding author

***Abstract*--Floods are natural events or phenomena that can cause environmental damage, loss of property, psychological effects, and death or casualties. One way to control flooding non-structurally is by mapping areas that are prone to flooding. This study builds a geographic-based information system to map flood-prone areas in Manado City using the K-Means Clustering algorithm. The main objective of this research is to identify and map areas with a high risk of flooding using spatial data. Slope, land cover type, soil type, water discharge (discharge), and rainfall are independent variables that will be used and processed using the K-Means Clustering algorithm. There are four clusters in the mapping results of flood-prone areas, namely: high vulnerability, medium vulnerability, low vulnerability, and not vulnerable. By using the K-Means method, the results obtained are Paal Dua and Wenang sub-districts are high-vulnerability groups, followed by Mapanget, Tuminting, and Singkil subdistricts with medium vulnerability groups. Tikala District is the only area with low vulnerability. Meanwhile, Bunaken, Sario, Wanea, and Malayayang sub-districts are areas that are not potentially prone to flooding.**

***Key Words*: Flood; K-Means Clustering; Geographic Information System.**

I. INTRODUCTION

Flooding is a natural occurrence that can inflict environmental damage, material loss, psychological impact on impacted individuals, and even death. Floods can last for an extended length of time, perhaps more than 24 hours. Furthermore, floods are defined as incidents in which the water level exceeds 40 cm and covers a vast area measuring more than 100 meters.

Rainfall, river water output, soil type, slope, and land cover are all interconnected factors that can produce flooding. Precipitation refers to the amount of rain that falls in a certain period of time, whereas river discharge refers to the highest volume of water that flows through a river in a given time period. Furthermore, the type of soil is related to its ability to absorb water, but the slope is tied to the area's topography, which can range from flat to mountainous. Flooding is more likely in places with gentle slopes than in locations with steep slopes. Land use in an area is connected to land cover.

Manado City flooded as an effect of heavy rain on January 27, 2023. Due to chronic rain, the Tondano River's water volume increased significantly, leading to flooding in various regions. This flood could not be avoided, and as it stands, many houses in numerous sub-districts, including Paal Dua, Tuminting, Sario, Wenang, and Singkil, were flooded. Given the possibility of flooding, it is critical to offer appropriate information about flood-prone areas. This will encourage people to be more cautious. As a result, it is vital to map regions in Manado City that have the potential to flood [1], so that we may identify areas in danger of flooding and those that are less prone to this disaster [2][3].

Previous research on K-Means for flood hazard mapping in Jombang [4] produced quite accurate results of 52.38%. Likewise, research by [5] resulted in the development of a technique for classifying flood-prone locations using color categorization. Green indicates places that are safe from floods, orange indicates moderately vulnerable areas, and red indicates extremely vulnerable areas. In addition, a recent study by [6] classified items depending on their distance from the center point, yielding three categories: susceptible regions, less vulnerable areas, and places not prone to floods. Then, a prior study by [7] entailed a digitizing process utilizing Arcview software to identify locations based on criteria for flooding hazards. The results are divided into three cluster levels, namely low cluster, medium cluster, and high cluster, for areas that are vulnerable to flooding.

The Geographic Information System (GIS) is a

EISSN 2654-802X ; PISSN 2985-4067
DOI: https://doi.org/10.56873/JITU.7.1.5403. SUBMITTED: DECEMBER 14, 2023; REVISED: MARCH 13, 2024; ACCEPTED: APRIL 5, 2024

platform for sharing information on location, space, and place, as well as changeable data [8][9][10]. GIS is utilized for mapping, and the K-means clustering algorithm is used to organize enormous volumes of data. This method aids in the identification and grouping of data depending on specific parameters. Aside from that, the K-Means Clustering approach will be represented using Geographic Information Systems (GIS) maps.

According to the explanation, the author sees a potential to deploy the K-means clustering approach in a geographic information system focused on mapping flood-prone zones in Manado City. Slope [10], land cover type [11], soil type [12] [13], water estimates (discharge) [14], and rainfall [15] are among the factors that will be employed. Aside from identifying and mapping flood-prone locations, another goal of this research is to segment risk zones using the K-Means Clustering method to classify places into risk zones based on geographic characteristics, hydrology, and other criteria. other pertinent matters. The risk zone segmentation data will be provided in map format, with four colors indicating the extent of vulnerability in each location. By segmenting risk zones, the government is able to take necessary action in risk mitigation and regional development planning.

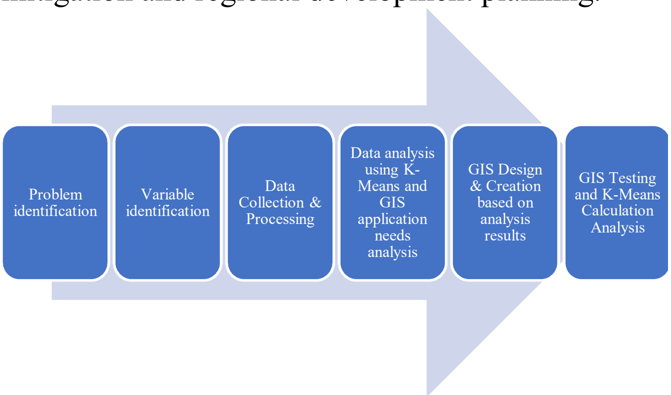

Fig. 1. Framework of Research

The system's computations must produce the same results as manual calculations during the analysis step. This GIS will be created as an online platform accessible over the internet [16], in order for its use to not be confined to the Manado City region. In this instance, it is accessible from anywhere, including outside the city. It is intended that the establishment of this system will assist the government in anticipating locations that may experience flooding and providing essential information to the public so that they are better aware of the risk of flooding in the future.

II. Methods

This research is quantitative and starts with data collection. Data was obtained from the Balai Wilayah Sungai Sulawesi I and Badan Perencanaan Pembangunan Daerah (Bappelitbangda) Manado. An interview was also conducted with one of the employees of the Balai Wilayah Sungai Sulawesi I to find out the flood mapping procedures carried out so far. The data obtained is in the form of slope, land cover, soil type, water discharge (discharge) and rainfall. Data were analyzed using the MSI (Method of Successive Interval) technique because this technique can convert ordinal data. The results from MSI are used for clustering using the K-Means method; prone to flooding or not. Meanwhile, RAD is used to create GIS applications. The RAD technique expedites and streamlines design. System development employs a repeatable procedure [17].

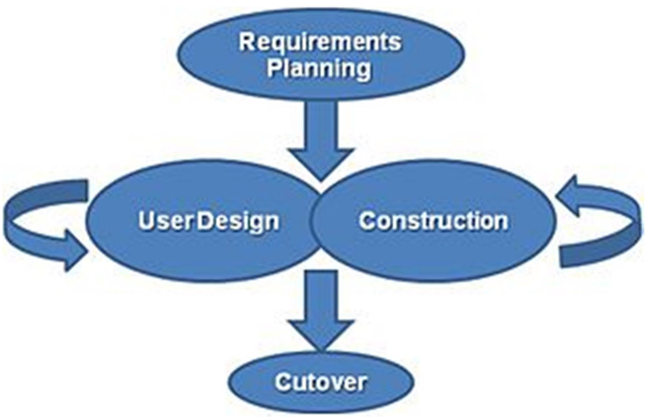

Fig. 2. Rapid Application Development Stages

1. Requirements Planning
   The initial stage focuses on locating suitable data sources. Data was gathered using direct research and observation methods, as well as interview approaches.
2. User Design
   This stage focuses on designing the application, producing an appealing appearance, identifying needs specifications, and developing a user-friendly system model.
3. Construction
   This stage focuses on the creation and development of applications. Programmers create features for applications. Users can

EISSN 2654-802X ; PISSN 2985-4067
DOI: https://doi.org/10.56873/JITU.7.1.5403. SUBMITTED: DECEMBER 14, 2023; REVISED: MARCH 13, 2024; ACCEPTED: APRIL 5, 2024

submit feedback or request modifications to the application.

4. Cutovers

Following the construction phase, this stage focuses on testing the application. This final stage consists of testing the function, interface, and overall program or application that has been produced, culminating in a new application that can run smoothly.

III. RESULT AND DISCUSSION

The data for this study were gathered through interviews with officials from Balai Wilayah Sungai Sulawesi I and Badan Perencanaan Pembangunan Daerah (Bappelitbangda). Soil type, slope level, land cover type, rainfall, and water discharge estimations are among the data collected.

A. Data Collection

Data collection is the collecting of information that describes each variable or item utilized in the study. Table I displays the results of the soil type data translation. Table II displays the results of the Slope Data. Table III displays the land cover data. Table IV explains the rainfall in each area in Manado over the last 11 years. Table V explains the debit data for 7 Pos Duga data Water in Manado City in the last 11 years (2011-2021).

TABLE I
Soil Type Data

No	Subdistrict	The average
1.	Bunaken	7,9567
2.	Mapanget	12,7432
3.	Tuminting	7,1267
4.	Singkil	7,1267
5.	Wenang	7,1267
6.	Paal Dua	9,7466
7.	Sario	7,1267
8.	Tikala	9,7466
9.	Wanea	9,7466
10.	Malalayang	17,4185

TABLE II
Slope Data

No	Subdistrict	The average
1.	Bunaken	4,3190
2.	Mapanget	3,9190
3.	Tuminting	3,9190
4.	Singkil	3,9190
5.	Wenang	3,9190
6.	Paal Dua	3,9190
7.	Sario	3,0154
8.	Tikala	3,9190
9.	Wanea	3,9190
10.	Malalayang	3,9190

TABLE III
Land Cover Data

No	Subdistrict	The average
1.	Bunaken	12,3312
2.	Mapanget	22,1099
3.	Tuminting	11,5078
4.	Singkil	9,7931
5.	Wenang	18,1439
6.	Paal Dua	19,1631
7.	Sario	14,5458
8.	Tikala	13,4711
9.	Wanea	20,7036
10.	Malalayang	19,5538

TABLE IV
Maximum Rainfall

Subdistrict	Maximum Rainfall	Final score
Bunaken	113,3000	113,3000
Mapanget	113,3000 126,7273	120,0136
Tuminting	113,3000	113,3000
Singkil	113,3000	113,3000
Wenang	131,2818 113,3000	122,2909
Paal Dua	131,2818 113,3000	122,2909
Sario	131,2818	131,2818
Tikala	131,2818 148,0727	139,6773
Wanea	131,2818 148,0727	139,6773
Malalayang	124,3455 131,2818	127,8136

TABLE V
Estimated Water

Subdistrict	Water Estimate (Discharge)	Final score
Bunaken	59,2300	59,2300
Mapanget	59,2300 151,8673	105,5487
Tuminting	59,2300 151,8673	105,5487
Singkil	151,8673	151,8673
Wenang	964,7311 151,8673	558,2992
Paal Dua	964,7311 151,8673	558,2992

EISSN 2654-802X ; PISSN 2985-4067
DOI: https://doi.org/10.56873/JITU.7.1.5403. SUBMITTED: DECEMBER 14, 2023; REVISED: MARCH 13, 2024; ACCEPTED: APRIL 5, 2024

Subdistrict	Water Estimate (Discharge)	Final score
Sario	29,7255 9,4628	19,5942
Tikala	964,7311 9,4628	487,0970
Wanea	29,7255 9,4628	19,5942
Malalayang	29,7255	29,7255

A. Calculation of K-Means Clustering

The results of the preceding investigation on numerous factors or characteristics resulted in a data collection of sub-district attributes. The data for District Parameters (Attributes) is shown in Table VI.

TABLE VI
District Data

Sub-district	Parameters (Attributes)				
	Type of soil	Slope	Land Cover	Rain-fall	Water Estimate (Discharge)
Bunaken	7,96	4,32	12,33	113,30	59,23
Mapanget	12,74	3,92	22,11	120,01	105,55
Tuminting	7,13	3,92	11,51	113,30	105,55
Singkil	7,13	3,92	9,79	113,30	151,87
Wenang	7,13	3,92	18,14	122,29	558,30
Paal Dua	9,75	3,92	19,16	122,29	558,30
Sario	7,13	3,02	14,55	131,28	19,59
Tikala	9,75	3,92	13,47	139,68	487,10
Wanea	9,75	3,92	20,70	139,68	19,59
Malalayang	17,42	3,92	19,55	127,81	29,73

The steps for computing K-Means Clustering are as follows:

1. Counting the number of Clusters.
   C1 is for High Susceptibility, C2 is for Medium Susceptibility, C3 is for Low Susceptibility, and C4 is for Unprone.
2. Selection of the first cluster center (cluster centroid) at random.

TABLE VII
Initial Cluster Center (Centroid)

C1	9,7466	3,9190	19,1631	122,2909	558,2992
C2	7,1267	3,9190	11,5078	113,3000	105,5487
C3	9,7466	3,9190	13,4711	139,6773	487,0970
C4	12,7432	3,9190	22,1099	120,0136	105,5487

3. Use the Euclidean Distance formula to calculate the distance between data and cluster centers, resulting in a table (matrix) of distances to clusters.

D (i,j) =

$$\sqrt{\left(x_{1_i}-x_{1_j}\right)^2+\left(x_{2_i}-x_{2_j}\right)^2+\ldots+\left(x_{k_i}-x_{k_j}\right)^2} \quad (1)$$

Bunaken District:

- Data distance from Bunaken District to centroid center C1:
  D (1,1) = 499,2113
- Data distance from Bunaken District to the C2 centroid center:
  D (1,2) = 46,3351
- Data distance from Bunaken District to the C3 centroid center:
  D (1,3) = 428,6847
- Data distance from Bunaken District to the C4 centroid center:
  D (1,4) = 48,0613

Calculations are carried out for each sub-district (carry out calculations for up to 10 sub-districts). Table VIII is the result of calculating the data distance to the centroid center in iteration 1.

TABLE VIII
Iteration 1

Subdistrict	Data Distance to Centroid Center in Iteration 1			
	C1	C2	C3	C4
Bunaken	499,2003	46,3351	428,6847	48,0540
Mapanget	452,7758	13,7485	382,1641	0,0000
Tuminting	452,9121	0,0000	382,4730	13,7485
Singkil	406,6477	46,3504	336,2961	48,7210
Wenang	2,8112	452,8884	73,4898	452,8085
Paal Dua	0,0000	452,9121	73,5149	452,7758
Sario	538,8070	87,8725	467,5876	87,2051
Tikala	73,5149	382,4730	0,0000	382,1641
Wanea	538,9878	90,4177	467,5588	88,2372
Malalayang	528,6584	78,2973	457,6301	76,4093

4. Group each object into the closest centroid to get cluster results.

TABLE IX
Cluster Grouping Results In Iteration 1

No	Subdistrict	*Cluster*
1.	Bunaken	C2
2.	Mapanget	C4
3.	Tuminting	C2
4.	Singkil	C2
5.	Wenang	C1
6.	Paal Dua	C1
7.	Sario	C4
8.	Tikala	C3

EISSN 2654-802X ; PISSN 2985-4067
DOI: https://doi.org/10.56873/JITU.7.1.5403. SUBMITTED: DECEMBER 14, 2023; REVISED: MARCH 13, 2024; ACCEPTED: APRIL 5, 2024

No	Subdistrict	*Cluster*
9.	Wanea	C4
10.	Malalayang	C4

5. Determining the new center point's position in order to carry out the following iteration.
6. Repeat step 3 until the new centroid location is the same as the old.

The computation method is repeated until the new centroid has the same value as the previous iteration's centroid. The average is derived from the data included in the same centroid to provide a new centroid. When the calculation tables for Iteration 2 and Iteration 3 are compared, the clusters in each sub-district have not changed (have been fixed), hence the computation concludes in Iteration 3. The final data acquired by clustering is shown in Tabel 10.

TABLE X
Final Results of K-Means Clustering Calculations

No	Subdistrict	*Cluster*	Status
1.	Bunaken	C4	Not vulnerable
2.	Mapanget	C2	Medium Susceptibility
3.	Tuminting	C2	Medium Susceptibility
4.	Singkil	C2	Medium Susceptibility
5.	Wenang	C1	High Vulnerability
6.	Paal Dua	C1	High Vulnerability
7.	Sario	C4	No potential
8.	Tikala	C3	Low Vulnerability
9.	Wanea	C4	No potential
10.	Malalayang	C4	No potential

B. *Design*

This section explains the flowchart design of the system that will be created. This flowchart depicts the stages involved in computing the K-Means Clustering technique. Fig. 3 shows the K-Means Clustering Flowchart in the program on the next page.

C. *Implementations*

Results of the implementation of the Geographic Information System Mapping Flood Prone Areas in Manado City Using the K-Means Clustering Method are shown in Fig. 4 - 7.

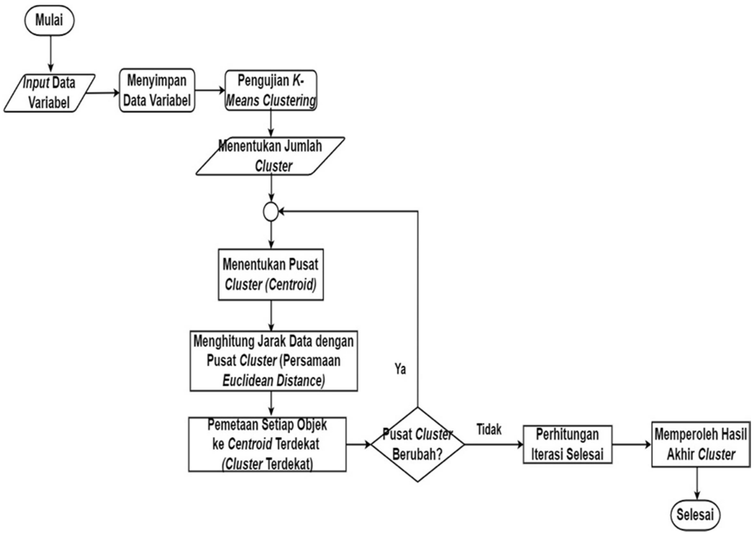

Fig. 3. Flowchart of K-Means Clustering in the Application

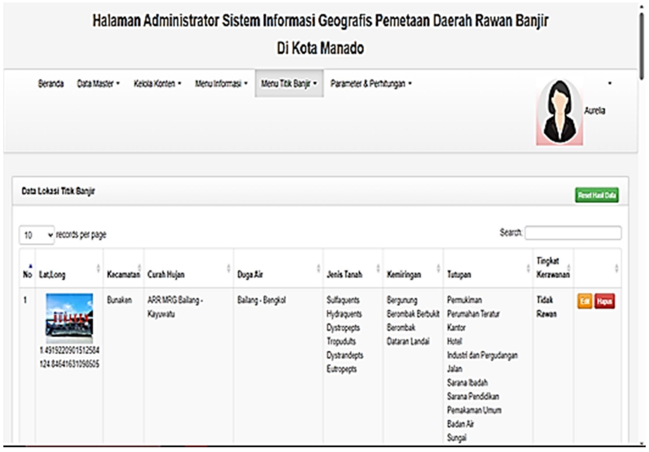

Fig. 4. Flood Point Location Data Page

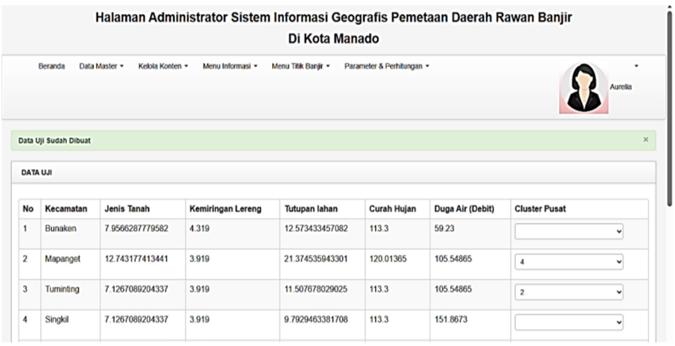

Fig. 5. K-Means Clustering Calculation Page

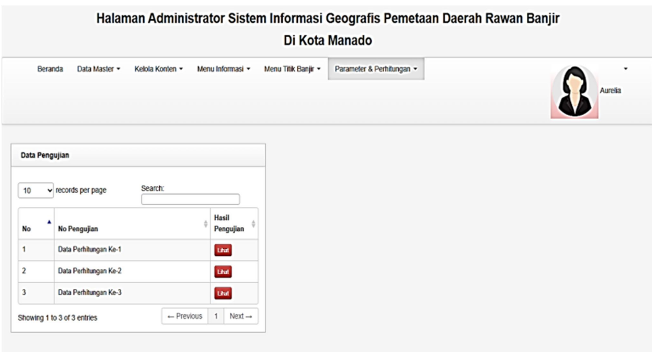

Fig. 6. K-Means Clustering Calculation Results Page

EISSN 2654-802X ; PISSN 2985-4067
DOI: https://doi.org/10.56873/JITU.7.1.5403. SUBMITTED: DECEMBER 14, 2023; REVISED: MARCH 13, 2024; ACCEPTED: APRIL 5, 2024

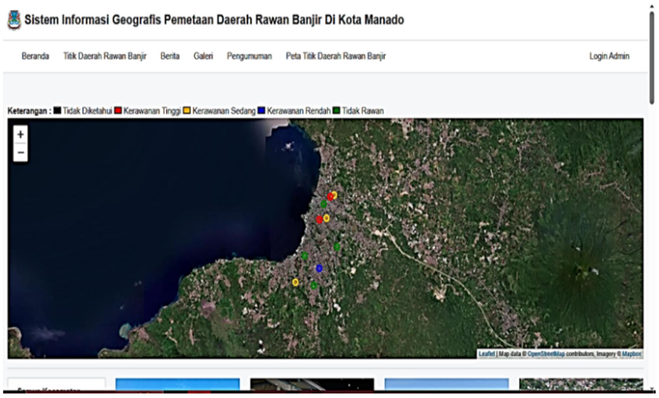

Fig. 7. Page Points for Flood Prone Areas

D. Testing

The system was thoroughly tested, including functionality, user interfaces, and any software or applications that have been developed. The goal of this test was to determine whether all features or functions comply with set standards and to identify any shortcomings that need to be remedied.

TABLE XI
Testing Results

Testing	Expected results	Actual Results
Select the "add data" button on the admin data page and enter the details of the data you want to add.	The data entered is added to the admin data list.	Admin data added successfully
Added 1 flood point location data on the flood point menu page.	The entered data is added to the flood point location data list.	Flood point data added successfully
Select the "create test data" button on the K-Means Clustering calculation page.	The entered data is added to the K-Means Clustering calculation test list.	Test data added successfully
Select a central cluster of more than 4 sub-districts	Cannot save the initial central centroid determination (because the vulnerability level is only 4 clusters).	The application gives a notification "Selected clusters are more than 4.
Select the "view" button for testing data from the K-Means Clustering calculation results.	Admin can see data from K-Means Clustering calculations and their iterations.	The application successfully displays the results of the K-Means Clustering calculation and its iterations.
Select a map of flood-prone areas.	Users can view a map page of flood-prone areas.	The application successfully displays a map page of flood-prone areas.

Testing of the functions and features of the GIS application that was built was successfully carried out and the application was free from errors, even the K-Means method was successfully applied to the application. By using the K-means method, the GIS application can depict flood-prone areas in the city of Manado easily and precisely. This is in accordance with the results of previous studies.

II. Conclusion

Based on the above explanation, it is possible to conclude that Geographic Information System has succeeded in displaying mapping of potentially flood-prone areas in Manado City using the K-Means Clustering method. Results of clustering of flood-prone areas using the K-means Clustering method, namely that the Paal Dua and Wenang areas are the most flood-prone areas, followed by Mapanget, Tuminting and Singkil as flood-prone areas. The Tikala area is not potentially prone to flooding along with Bunaken, Sario, Wanea and Malayayang.

For further research, rainfall data should use data from BMKG Manado City. Additional attributes or variables that cause flooding can be added according to the conditions of an area. This research can use the Meta-Heuristic algorithm to determine the centroid center so that clustering results are better.

III. References

[1] I. Kairupan, Z. Y. Huang, H. C. Chang, and C. W. Chang, "Emergency Navigation and Alarm with Flooding Models - A Real Case Study Of Manado City," in *2016 International Conference on Communication Problem-Solving, ICCP 2016*, Taiwan: IEEE, 2016, pp. 1–2. Doi: 10.1109/Iccps.2016.7751120.

[2] H. Rakuasa, J. K. Helwend, and D. A. Sihasale, "Pemetaan Daerah Rawan Banjir di Kota Ambon Menggunakan Sistem Informasi Geografis," *Jurnal Geografi, Media Informasi Pengembangan dan Profesi Kegeografian*, Vol. 19, No. 2, 2022, Doi: 10.15294/Jg.V19i2.34240.

EISSN 2654-802X ; PISSN 2985-4067
DOI: https://doi.org/10.56873/JITU.7.1.5403. SUBMITTED: DECEMBER 14, 2023; REVISED: MARCH 13, 2024; ACCEPTED: APRIL 5, 2024

[3] P. F. Opit, I. Y. Kairupan, and F. M. Rusuh, "A Milp Model for Water Level Sensor Placement with Multi-Sensor and Multi-Disaster Areas," *J. Tek. Ind.*, Vol. 22, No. 2, 2021, Doi: 10.22219/Jtiumm.Vol 22. No 2.185-195.

[4] R. A. Kurniawan. "Sistem Informasi Geografis Pemetaan Daerah Potensi Rawan Banjir dengan Metode K-Means Clustering (Studi Kasus: Kabupaten Jombang)". Skripsi. Universitas Bhayangkara Surabaya. 2023.

[5] I. Faturohim and A. Baita, "Penerapan Algoritma K-Means Clustering untuk Mengelompokkan Kerawanan Rob di Daerah Pekalongan," *Juti "Jurnal Teknol. Informasi,"* Vol. 1, No. 1, 2022, Doi: 10.26798/Juti.V1i1.624.

[6] R. Aprillina, A. Nurlifa, A. Haryoko, R. E. Putri, and A. N. Rosalita, "Clustering Daerah Rawan Banjir di Kabupaten Tuban dengan K-Means Disertai Visualisasi," *Curtina Comput. Sci. Or Inform. J.*, Vol. 1, No. 1, Pp. 40–51, 2022.

[7] N. N. Adisanjaya, A. Kusuma, and I. G. N. M. Nugraha, "Pemetaan Zonasi Daerah Rawan Banjir di Denpasar Bali Dengan Metode K-Means Clustering," *J. Media Sains*, 2021.

[8] A. F. Aslamsyah, Y. A. Pranoto, and R. P. P, "Sistem Informasi Geografis Daerah Hortikultura (Sayuran) Kabupaten Malang," *Jati (Jurnal Mhs. Tek. Inform.*, Vol. 4, No. 2, 2020.

[9] K. P. Sinaga and M. S. Yang, "Unsupervised K-Means Clustering Algorithm," *IEEE Access*, Vol. 8, 2020, Doi: 10.1109/Access.2020.2988796.

[10] B. Lund and J. Ma, "A Review of Cluster Analysis Techniques and Their Uses in Library and Information Science Research: K-Means and K-Medoids Clustering," *Perform. Meas. Metrics*, Vol. 22, No. 3, 2021, Doi: 10.1108/Pmm-05-2021-0026.

[11] N. P. G. Qur'ani, D. Harisuseno, and J. S. Fidari, "Studi Pengaruh Kemiringan Lereng terhadap Laju Infiltrasi," *J. Teknol. dan Rekayasa Sumber Daya Air*, Vol. 2, No. 1, 2022, Doi: 10.21776/Ub.Jtresda.2022.002.01.19.

[12] A. Yusuf, D. I. Kusumastuti, and E. P. Wahono, "Pengaruh Tutupan Lahan terhadap Base Flow Index Das Way Seputih Provinsi Lampung," *Siklus J. Tek. Sipil*, Vol. 7, No. 2, 2021, Doi: 10.31849/Siklus.V7i2.7323.

[13] A. Anwari and M. Makruf, "Pemetaan Wilayah Rawan Bahaya Banjir di Kabupaten Pamekasan Berbasis Sistem Informasi Geografis (Sig)," *Netw. Eng. Res. Oper.*, Vol. 4, No. 2, Pp. 117–123, 2019, Doi: 10.21107/Nero.V4i2.127.

[14] G. Genesiska, M. Mulyono, and A. Intan Yufantari, "Pengaruh Jenis Tanah terhadap Pertumbuhan dan Hasil Tanaman Jagung (Zea Mays L.) Varietas Pulut Sulawesi," *Plantropica J. Agric. Sci.*, Vol. 5, No. 2, 2020, Doi: 10.21776/Ub.Jpt.2020.005.2.2.

[15] Sarmidi and Sidik Ibnu Rahmat, "Sistem Peringatan Dini Banjir Menggunakan Sensor Ultrasonik Berbasis Arduino Uno," *J. Manaj. dan Tek. Inform.*, Vol. 02, No. 01, pp. 181–190, 2018.

[16] W. S. Aji, "Simulasi Dengan Multi Agent Prediksi Banjir Berdasarkan Intensitas Hujan Menggunakan Particle Swarm Optimization," *J. Teknol. Inf.*, Vol. 5, No. 2, 2019.

[17] T. S. Pratiwi and H. C. Chotimah, "Aktivitas Diplomasi Digital dalam Manajemen Bencana: Studi Kasus di Daerah Istimewa Yogyakarta, Indonesia dan Fukushima, Jepang," *J. Stud. Diplomasi dan Keamanan*, Vol. 13, No. 1, 2021, Doi: 10.31315/Jsdk.V13i1.4367.

[18] D. Hariyanto, et al., "Implementasi Metode RAD pada Sistem Informasi Perpustakaan", Jurnal Penelitian Ilmu dan Teknologi Komputer (JUPITER), Vol. 13, No. 1, pp. 110–117, 2021.