InJAST
Volume 5, Number 1, April 2024
Pages: 6-13

P - ISSN: 2722 - 0133
E - ISSN: 2722 - 0141
DOI: 10.33751/injast.v5i1.10259

Analysis of flood-prone areas in DKI Jakarta Province using Clustering Method

RANDY DAFFA ADITYA*, MUHAMMAD ABDUL AZIZ HABIBI

Politeknik Statistika STISS, Jl. Otto Iskandardinata No.64C-1, Jatinegara, Kota Jakarta Timur, Jakarta 13330, Indonesia

Corresponding author: *212112306@stis.ac.id*

Submitted 10 July 2024; Accepted 16 August 2024

ABSTRACT

The objective of this research is to ascertain the patterns and organization of flood-affected areas in Jakarta. The dataset of flood incidents in the DKI Jakarta Province in 2020 served as the data source for this study. The research employed three methods: K-Means, K-Medoid, and Hierarchical Clustering. Of these, Hierarchical Clustering produced the best grouping in comparison to the other methods. The findings of the study show that the flood-affected areas in DKI Jakarta are classified into three groups: safe (cluster 1), moderate (cluster 2), and vulnerable (cluster 3). The districts of Cengkareng, Jatinegara, and Pulogadung are among the vulnerable areas.

ABSTRAK

Tujuan penelitian ini adalah untuk mengetahui pola dan penataan wilayah terdampak banjir di Jakarta. Dataset kejadian banjir di Provinsi DKI Jakarta tahun 2020 dijadikan sebagai sumber data penelitian ini. Penelitian ini menggunakan tiga metode: K-Means, K-Medoid, dan Hierarchical Clustering. Dari ketiga metode tersebut, Hierarchical Clustering menghasilkan pengelompokan terbaik dibandingkan dengan metode lainnya. Temuan penelitian menunjukkan bahwa wilayah terdampak banjir di DKI Jakarta diklasifikasikan menjadi tiga kelompok: aman (kluster 1), sedang (kluster 2), dan rentan (kluster 3). Kecamatan Cengkareng, Jatinegara, dan Pulogadung termasuk wilayah yang rentan.

Keywords: *clustering, flood, Jakarta, vunerable*

INTRODUCTION

Flooding is a natural phenomenon that can occur throughout the globe. Settlements in floodplain areas are the main cause of flood damage (Istiadi & Priatna, 2021). Jakarta is Indonesia's capital city, and it is dealing with a number of issues. Poverty, inequality, and environmental concerns are all issues that both the government and society must address. Flooding is an environmental concern that has received a lot of attention. Major floods in 2002, 2007, 2013, and 2014 resulted in direct and indirect economic damages worth trillions of rupiah (Bappenas, 2007; Ward et al., 2013). Although flooding is not a new phenomenon, the extent of the damage has escalated considerably over the last decade.

Based on Figure 1 above, it is evident that each year, the number of neighborhoods (RW) affected by floods in Jakarta generally ranges between 200 and 300. These numbers tend to increase each year, with the exception of 2020. In 2020, the number of affected neighborhoods was exceptionally high due to extremely high rainfall (377 mm/day on January 1, 2020), causing widespread flooding across Jakarta. Nevertheless, the increasing number of neighborhoods affected by floods indicates that the damage caused by floods will continue to escalate each year.

This increase is associated with a myriad of factors, both physical and socio-economic. Physical factors include land subsidence, low drainage or storage capacity in Jakarta's rivers and canals due to blockage by waste and sediment eroded from upstream, as well as climate change. Socio-economic factors involve rapid population growth and changes in land use, leading to the development of economic assets in flood-prone areas (Budiyono, 2016). Climate change is also a major contributor to the frequent flooding in Jakarta (Douglass *in* Lyons, 2015; Fuchs in Salim et al., 2019). Climate change disrupts the schedule and intensity of rainfall, causing high sea levels and leading to floods and land subsidence (Padawangi & Douglass, 2015). This situation is exacerbated by Jakarta's poor drainage system and the

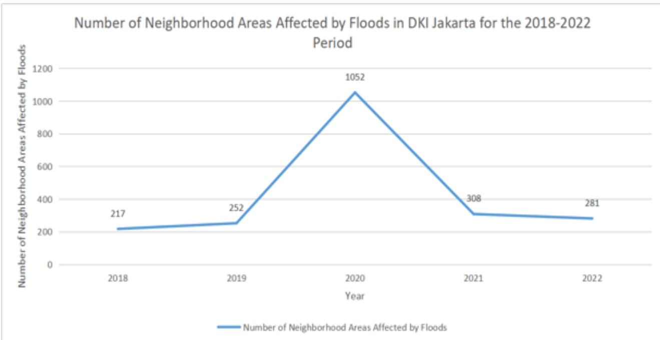

Figure 1. Graph of the number of neighborhood units (RW) affected by floods in DKI Jakarta for the period 2018-2022.

limited natural water absorption areas, preventing the efficient flow of water from other areas, rainwater, and seawater, resulting in overflowing drainage systems and rivers that cause floods (Lyon, 2015). Poor drainage maintenance combined with climate change causes Jakarta to experience floods annually (Napier, 2021; Padawangi & Douglass, 2015; Salim et al., 2019).

In addition to physical factors, socio-economic factors also influence flooding in Jakarta (Budiyono, 2016). This is partly due to Jakarta's status as the capital and economic center of Indonesia. The influx of people moving to or living in Jakarta has led to the construction of numerous buildings (Padawangi & Douglass, 2015). The proliferation of buildings has replaced many water absorption areas, resulting in an ineffective water absorption and drainage system in Jakarta (Padawangi & Douglass, 2015). This has made Jakarta increasingly vulnerable to flooding. The situation is further aggravated by the numerous illegal buildings along riverbanks and the significant amount of waste clogging the city's drainage system (Padawangi & Douglass, 2015).

Given these circumstances, flood management and prevention have become crucial programs in Jakarta. However, if the government does not prioritize the areas that need immediate flood management, their efforts may not be effective. Therefore, identifying the neighborhoods frequently affected by floods is essential to facilitate disaster mitigation and find solutions to address flooding. By knowing the flood-prone areas, those regions can better prepare themselves, thereby minimizing casualties and damage (Duykers et al., 2023).

Based on the explanation above, this research aims to identify the patterns and clustering of floods in Jakarta. Flood clustering is necessary because floods generally inundate multiple areas simultaneously, implying that some regions may be more vulnerable to flooding than others. By clustering flood-affected areas in Jakarta, it is hoped that disaster mitigation processes and flood management solutions can be implemented more easily and efficiently.

METHODS

Data Source

This research uses secondary data obtained from the website of data.jakarta.go.id, provided by the Provincial Government of DKI Jakarta. The data on this website is a compilation from various agencies such as the Badan Pusat Statistik (BPS), Badan Nasional Penanggulangan Bencana (BNPB), Dinas Kesehatan Provinsi DKI Jakarta, and other agencies. The dataset consists of 902 observations and includes attributes such as the administrative area where the flooding occurred, neighborhood units (RW), the number of RW, neighborhood associations (RT), households (KK) affected by the flood, water level, the number of affected households, number of refugees, number of fatalities, and the flood period. The collected data is cross-sectional with the unit of analysis being sub-districts in DKI Jakarta in the year 2020.

Methodology

The method used in this research is K-Means Clustering to discover and group data, enabling better insights into flooding in the DKI Jakarta area. K-Means clustering is one of the classic methods in unsupervised learning. This method is used to divide n units of observation into k groups or clusters so that each cluster has an average value as close as possible (Alashwal et al., 2019). The notion of similarity can be expressed in very different ways, depending on the research objectives, specific domain assumptions, and prior knowledge about the issue (Grira et al., 2004).

As the name suggests, K-Means Clustering requires prior information about how many clusters will be generated according to previous information or research (Sinaga & Yang, 2020). The value of k will then be initialized into the dataset of n observations to generate a number of clusters along with their members and the average values of each cluster. In summary, the following is the procedure for using K-Means Clustering (Alashwal et al., 2019).

1. Initialize k cluster centers.
2. Calculate the distance between each observation and the cluster center observation points.
3. Assign all points to the cluster whose center is at the minimum distance from all cluster centers.
4. Recalculate the positions of the k centers as the mean of the clusters.
5. Recalculate the distance between each data point and the newly calculated centers. Repeat steps 3 and 4 until all data points are assigned to the samecluster (data points also do not move).

The K-Means Clustering method was chosen because it has been proven effective in grouping data into specific clusters or groups according to the research objectives. K-means clustering has been extensively studied with various extensions in the literature and applied in various substantive fields (Sinaga & Yang, 2020). The advantage of this method is its ability to learn data in an unsupervised manner, meaning data without target variables (Alashwal et al., 2019). This is highly suitable for this research on flood clustering in the DKI Jakarta area. However, this method also has a drawback, namely, its heavy reliance on the value of k initialized by the researcher. Therefore, researchers must be cautious in determining the number of clusters to ensure optimal clustering results.

In the use of K-Means Clustering, other methods such as K-Medoid and Hierarchical Clustering are sometimes necessary. This is because K-Means Clustering is highly affected by outliers that can disrupt the data (Kaur et al., 2014). This is due to the calculation method of K-Means Clustering itself, which calculates based on the distance between the values of elements in a group and the group's average (Medellu & Nugaraha *in* Herman et al., 2022). Therefore, the K-Medoid method emerged as an alternative clustering method by selecting the most central value within each cluster (Kaur et al., 2014). The process works as follows: from numerous data points, a number of medoids are selected randomly, where groups are formed with elements most similar to the representative values. After these clusters are formed, new medoids are selected that better represent the formed groups. This continues until no medoids change their positions (Vishwakarma et al., *in* Herman et al., 2022). The K-Medoid method provides results that are more resilient to outliers and more robust, although the algorithm becomes more complex, and the number of clusters formed still heavily depends on the researcher's decisions (Kaur et al., 2014; Herman et al., 2022).

Hierarchical Clustering is a clustering method that involves either merging or splitting existing groups and specifying the order of cluster merging or splitting (Shetty & Singh, 2021). A tree or dendrogram is used to display hierarchical clusters. Hierarchical clustering can be performed in two ways: bottom-up (Agglomerative Hierarchical Clustering) or top-down (Divisive Hierarchical Clustering). Large clusters are divided into smaller clusters, and small clusters from large clusters are merged into one (Shetty & Singh, 2021). The steps to form top-down clustering are:

1. Start with all data points as individual clusters.
2. After each iteration, remove the "outlier" from the least cohesive cluster.
3. Repeat the second step, stopping when each example is in its own single cluster.

The steps of agglomerative (bottom-up) clustering formation are:

1. Begin by considering each data point as its own single cluster.
2. After each iteration of Euclidean distance calculation, merge two clusters with the minimum distance.

Repeat the second step, stopping when there is one cluster containing all examples.

RESULT AND DISCUSSION

Descriptive Statistics

Here is a summary of the description regarding flood impact data in DKI Jakarta Province used in the study.

Based on Table 1, it is known that on average there are seven RW (neighborhood units) affected by floods, with a minimum of one RW and a maximum of 15 RW, and a resulting standard deviation of 3.527. Furthermore, there are on average 14.95 RT (household units) affected, with a minimum of 1 RT and a maximum of 46 RT. The standard deviation of the number of affected RT is 11.889. Regarding the variable of the number of affected households (KK), on average there are 394.41 households affected by floods, with a minimum of 0 and a maximum of 3762 households. The standard deviation value is 727.334. In terms of the variable of the number of affected individuals, there are on average 1414 individuals affected by floods, with a minimum of 0 and a maximum of 13450 individuals, and the standard deviation is 2446.895. As for the water level, the average height is 87.05 cm, with a minimum of 20 cm and a maximum of 180 cm. The standard deviation value is 40.35. For the duration of inundation, floods typically last for an average of 0.3182 days, with a minimum of 0 days and a maximum of 2 days. The standard deviation is 0.561. Lastly, for the variable of the number of minor injuries, on average there are 811.79 individuals sustaining minor injuries, with a minimum of 0 and a maximum of 4461 individuals. The standard deviation is 1036.664.

Table 1. Summary of descriptive statistics.

Variables	Mean	Minimum	Maximum	Standard Deviation
Number of Neighborhood Association (RW) affected	7	1	15	3.527
Number of Community Unit (RT) affected	14.95	1	46	11.889
Number of Household (KK) affected	394.41	0	3762	727.334
Number of Individuals affected	1414	0	13450	2446.895
Water Level (cm)	87.05	20	180	40.350
Inundation Duration (day)	0.3182	0	2	0.561
Number of Minor Injuries	811.79	0	4461	1036.664

Data Preprocessing

Data preprocessing is a crucial phase in data analysis as it has the potential to influence the final results of the analysis or models built. This process can be customized and depends on the type of data and the objectives of the analysis or models to be achieved. The stages of data preprocessing include several steps, namely data cleaning, data transformation, and data reduction.

Before processing the data, data cleaning is performed first. Data cleaning is a stage where data is filtered to ensure its accuracy, consistency, completeness, and readiness for analysis. In this study, data cleaning focuses on checking for missing or empty values (missing values) in the dataset. After that, a check for missing values is performed. The result is that there are no missing values in the dataset.

After the completion of data cleaning, the process continues with data transformation. This data transformation will change or modify the data from its original format to another format for a specific purpose. Data transformation is done to ensure that the data to be analyzed or processed meets the existing needs. This change is made because there are changes in units and dimensions in the data, so standardization with z-score is necessary.

The final stage of data preprocessing is data reduction. Data reduction is the process of reducing the amount of irrelevant, excessive, or redundant data in a dataset. The goal is to reduce unnecessary analysis complexity, save time and resources in data processing, and improve the efficiency and quality of data analysis. In this data reduction process, KMO (Kaiser-Meyer-Olkin) and PCA (Principal Component Analysis) analyses will be conducted. KMO is the abbreviation for the Kaiser-Meyer-Olkin Measure of Sampling Adequacy, which is used to assess the adequacy of the data sample used in factor analysis or PCA. PCA, on the other hand, transforms linear combinations of the original attributes into several principal components, selected based on their contribution to the variation in the data. These components provide a more concise representation of the original data, combining information from several original attributes that contribute most to the variation in the data. By using these reduced attributes, data analysis can be performed more efficiently while still retaining the important information represented by these principal components.

In this dataset, the variables for the number of deaths, number of missing persons, number of severe injuries, and the value of losses all have a value of 0, so it was decided to remove them. After that, KMO analysis is performed to determine its value. If the MSA value is more than 0.5, then the attribute can be retained. Whereas if it is less than 0.5, then the attribute will be removed. Then from the output, the overall MSA value obtained is 0.74, while the MSA value for each retained variable is listed in Table 2.

Table 2. MSA Score for each variables.

Variables	MSA Score
Number of Neighborhood Association (RW) affected	0.82
Number of Community Unit (RT) affected	0.83
Number of Household (KK) affected	0.63
Number of Individuals affected	0.64
Water Level (cm)	0.90
Inundation Duration (day)	0.67
Number of Minor Injuries	0.89

K-Medoid Clustering Method

The optimal number of clusters to be used in the K-medoid method can be determined using Elbow Plot visualization. Additionally, the determination of the optimal number of clusters can also be observed using Connectivity, Silhouette, and Dunn indices. The Elbow Plot and indices can be seen in Figure 2 and Table 3.

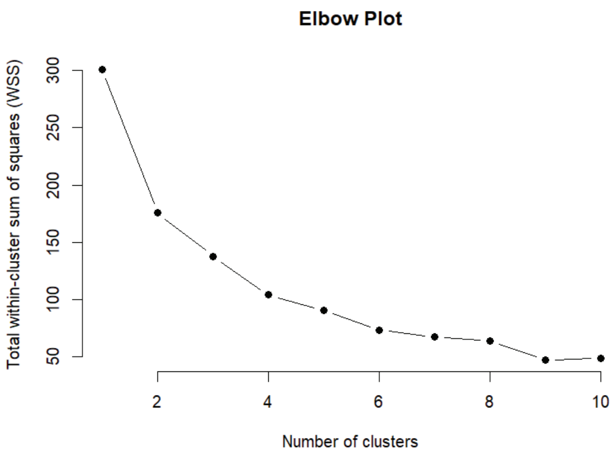

Figure 2. K-Medoid method's elbow plot.

Table 3. K-Medoid validity index value.

Validation Measures	k=3	k=4	k=5
Connectivity	15.3218	19.3032	20.8365
Dunn	0.1342	0.1601	0.2273
Silhouette	0.2793	0.3056	0.3098

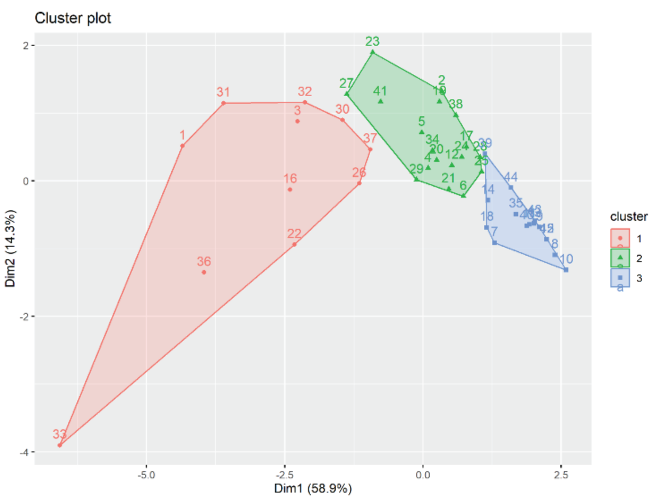

Figure 3. Cluster Plot using K-Medoid algorithm.

Based on Table 3, the smallest Validation Measure Connectivity value is found in cluster 3, thus it can be concluded that the optimal number of clusters in the K-Medoid method to be used is 3 clusters. Visualization of the clustering results with k = 3 is shown in Figure 3 and Figure 4.

Based on the choropleth map in Figure 4, the distribution of flood-prone area clusters is divided into 3 groups. The first cluster consists of 11 districts, the second cluster consists of 18 districts, and the third cluster consists of 15 districts.

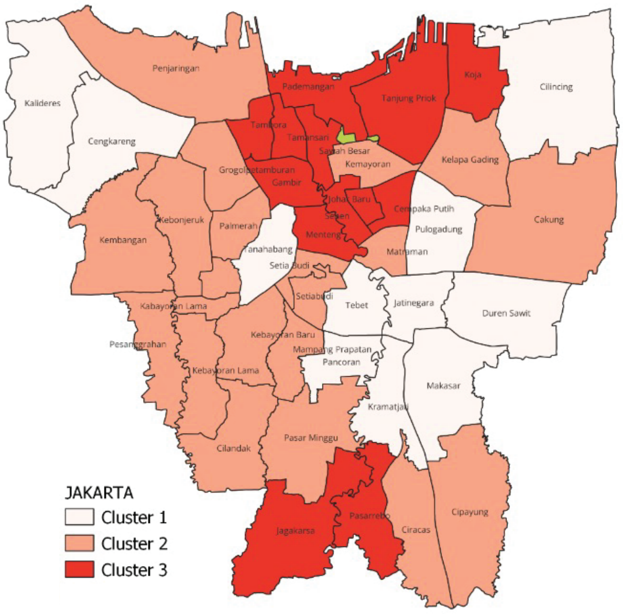

Figure 4. Mapping of flood-prone cluster regions in Jakarta using the K-Medoid method.

K-Means Clustering Method

The optimal number of clusters to be used in the K-means method can be determined using Elbow Plot visualization. Additionally, the determination of the optimal number of clusters can also be observed using Connectivity, Silhouette, and Dunn indices. The Elbow Plot and indices can be seen in Figure 5 and Table 4 below.

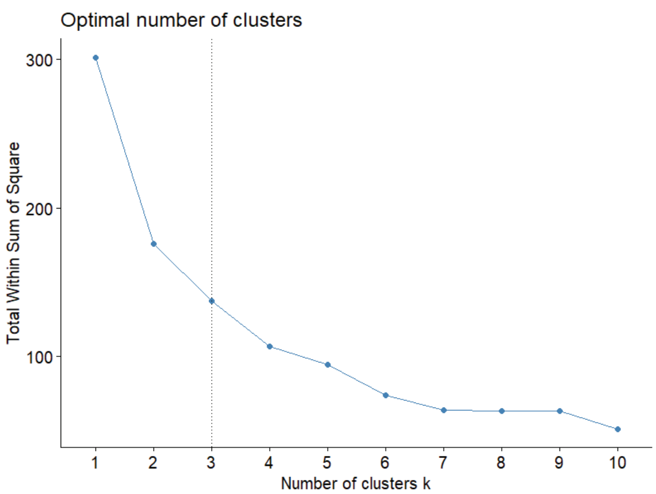

Figure 5. K-Mean Method's elbow plot.

Table 4. K-Means validity index value.

Validation Measures	k=3	k=4	k=5
Connectivity	14.6790	20.8071	21.9869
Dunn	0.3342	0.3542	0.3976
Silhouette	0.4005	0.3671	0.3573

Based on Table 4, the smallest value of Validation Measure Connectivity and the largest Dunn value are found in cluster 3, indicating that the optimal number of clusters for the K-Mean method to be used is 3 clusters. The visualization of the clustering results with k = 3 is shown in Figure 6 and Figure 7.

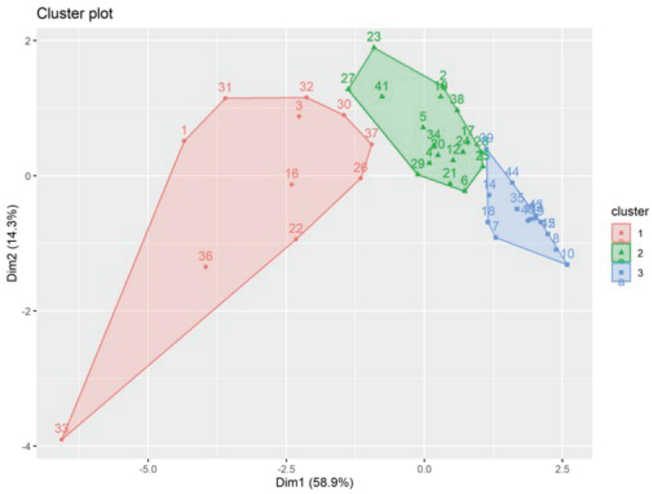

Figure 6. Cluster plot using K-Means algorithm.

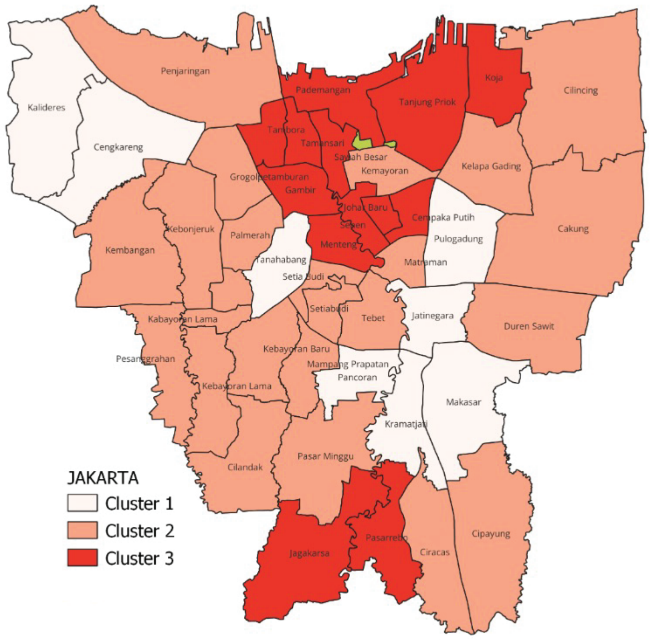

Figure 7. Mapping of flood-prone cluster regions in Jakarta using the K-Means method.

Based on the choropleth map using the K-Means method in Figure 7, the distribution of flood-prone cluster areas is divided into 3 groups. The first cluster consists of 8 districts, the second cluster consists of 21 districts, and the third cluster consists of 15 districts.

Hierarchical Method Clustering

In hierarchical clustering, the initial step is to select the best model. Model selection is done by examining the cophenetic correlation values of each available method.

The model with the highest cophenetic correlation value will be considered the best. Information regarding the cophenetic correlation values can be found in Table 5 below.

Table 5. Cophenetic correlation coefficient.

Methods	Correlation Coefficient
Single	0.8821
Complete	0.7999
Average	**0.9123**
Ward	0.6259

Based on the evaluation of the cophenetic correlation coefficient, the average linkage method stands out with the highest correlation compared to other alternatives, thus being selected for the clustering process. The determination of the number of clusters in the average linkage method is done through cluster validity testing first. In this study, internal validity testing is conducted using the Connectivity, Silhouette, and Dunn indices. The optimal number of clusters can be identified from the Silhouette and Dunn index values approaching 1, as well as the Connectivity value decreasing. Detailed values of internal validity indices are listed in Table 6.

Table 5. Cophenetic correlation coefficient.

Validation Measures	k=3	k=4	k=5
Connectivity	9.6341	10.6520	19.8119
Dunn	0.4275	0.4275	0.3356
Silhouette	0.4501	0.4128	0.3564

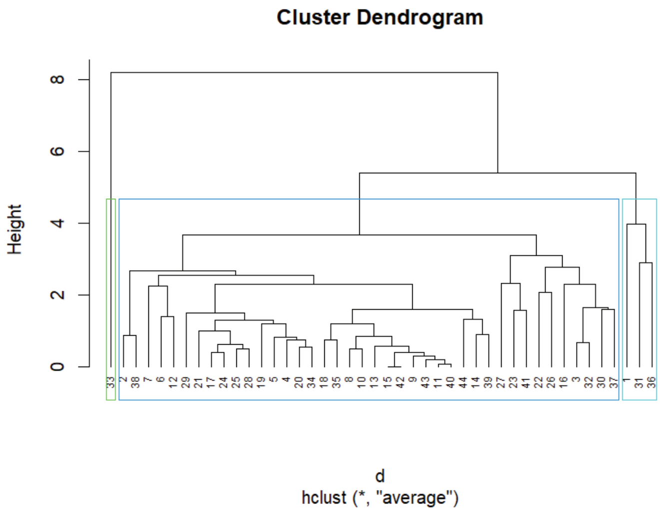

Figure 8. Dendrogram cluster using the average linkage method.

Based on Table 6, the smallest value of the Validation Measure Connectivity, and the largest values of Dunn and Silhouette, are found in cluster 3. Therefore, it can be concluded that the optimal number of clusters in the average linkage method to be used is 3 clusters. Visualization of the clustering results with k = 3 can be seen in Figure 8 and Figure 9.

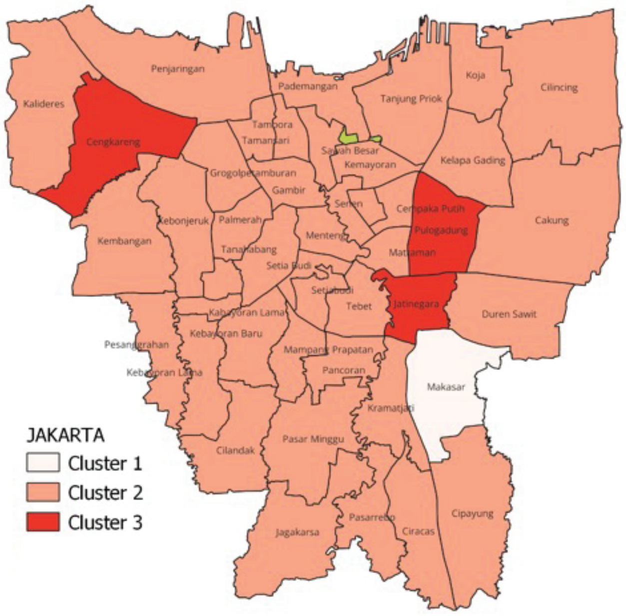

Figure 9. Mapping of flood-prone cluster regions in Jakarta using the average linkage method.

Based on the choropleth map using the Average Linkage method in Figure 9, the distribution of flood-prone cluster areas is divided into 3 groups. The first cluster consists of 1 district, the second cluster consists of 40 districts, and the third cluster consists of 3 districts.

Comparison between K-Medoid, K-Means, and Average Linkage Clustering Method

After conducting the clustering process using three different methods, the three methods will be compared to see which one is the best to use. To compare the three methods, the criteria are based on the Connectivity index, Dunn index, and Silhouette index. The best method is determined by which method has the smallest Connectivity value and the largest Dunn and Silhouette values. The comparison of the three clustering methods can be seen in the following Table 7.

Table 7. Validation index comparison for three utilized clustering method.

	Clustering Methods (k=3)		
Validation Measures	K-Medoid	K-Means	**Average Linkage**
Connectivity	15.3218	14.6790	**9.6341**
Dunn	0.1342	0.3342	**0.4275**
Silhouette	0.2793	0.4005	**0.4501**

Based on Table 7 above, it can be observed that the Average Linkage method yields the most optimal values of Connectivity, Dunn, and Silhouette indices compared to the K-Medoid and K-Means methods. Subsequently, based on the clustering results, the characteristics of each cluster are obtained based on the variables initially used in Table 8.

Table 8. Characteristic comparison for each cluster.

Variables	Cluster 1	Cluster 2	Cluster 3
Number of Neighborhood Association (RW) affected	10	6.6	**12.33**
Number of Community Unit (RT) affected	26	13.2	**34.667**
Number of Household (KK) affected	3762	200.125	**1862.333**
Number of Individuals affected	13450	784.15	**5800.667**
Water Level (cm)	115	82.85	**133.667**
Inundation Duration (day)	**2**	0.275	**0.333**
Number of Minor Injuries	2738	610.8	**2849.667**

Based on the comparison in table 8, cluster 3 exhibits high characteristics in terms of the variables of the number of affected neighborhoods (RW), the number of affected communities (RT), the number of affected households (KK), the number of affected individuals, water level, and the number of minor injuries. Meanwhile, the area with the longest duration of inundation is found in cluster 1. The grouping of areas based on their clusters is presented in Table 9.

Table 9. Inter-cluster area distribution.

Cluster Type	City	District
1 (Safe)	East Jakarta	Makassar
2 (Moderate)	West Jakarta	Grogol Petamburan, Kalideres, Kebon Jeruk, Kembangan, Palmerah, Taman Sari, Tambora
	Central Jakarta	Cempaka Putih, Gambir, Johar Baru, Kemayoran, Menteng, Sawah Besar, Senen, Tanah Abang
	South Jakarta	Cilandak, Jagakarsa, Kebayoran Baru, Kebayoran Lama, Mampang Prapatan, Pancoran, Pasar Minggu, Pesanggrahan, Setiabudi, Tebet
	East Jakarta	Cakung, Cipayung, Ciracas, Duren Sawit, Kramat Jati, Matraman, Pasar Rebo
	North Jakarta	Cilincing, Kelapa Gading, Koja, Pademangan, Penjaringan, Tanjung Priok
	Seribu Islands	Kepulauan Seribu Selatan, Kepulauan Seribu Utara
3 (Vunerable)	West Jakarta	Cengkareng
	East Jakarta	Jatinegara, Pulogadung

CONCLUSION

According to the average linkage method, flood-affected areas in DKI Jakarta are divided into 3 clusters: safe (cluster 1), moderate (cluster 2), and vulnerable (cluster 3). Areas categorized as vulnerable include the districts of Cengkareng, Jatinegara, and Pulogadung. The results of this study can be used for evaluation by the government, especially the Jakarta Provincial Government, to improve drainage infrastructure and related water channels to minimize losses and reduce casualties in the future.

REFERENCES

Abidin, H. Z., Andreas, H., Djaja, R., Darmawan, D., & Gamal, M. (2008). Land subsidence characteristics of Jakarta between 1997 and 2005, as estimated using GPS surveys. *GPS Solutions, 12: 23-32*.

Alashwal, H., El Halaby, M., Crouse, J. J., Abdalla, A., & Moustafa, A. A. (2019). The application of unsupervised clustering methods to Alzheimer's disease. *Frontiers in Computational Neuroscience, 13: 31*.

Alfahmi, F., Hakim, O. S., Dewi, R. C., & Khaerima, A. (2019). Utilization of data mining classification techniques to identify the effect of Madden-Julian Oscillation on increasing sea wave height over East Java Waters. *IOP Conference Series: Earth and Environmental Science, 399(1): 012062*.

Budiyono, Y., Aerts, J. C., Tollenaar, D., & Ward, P. J. (2016). River flood risk in Jakarta under scenarios of future change. *Natural Hazards and Earth System Sciences, 16(3): 757-774*.

Duykers, J. L., Ardana, K., Yulistiani, R., Irwansyah, E., & Fitrianah, D. (2023). Identifying Factors for Supporting Early Warning Flood using Clustering Approach and Geo-Spatial Analysis. *Procedia Computer Science, 227: 540-547*.

Erraguntla, M. et al. (2014). Data Mining to Improve Safety of Blood Donation Process. 47th Hawaii International Conference on System Sciences, Waikoloa, HI, USA, 2014, pp. 789-795, doi: 10.1109/HICSS.2014.105.

Grira, N., Crucianu, M., & Boujemaa, N. (2004). Unsupervised and semi-supervised clustering: a brief survey. *A Review of Machine Learning Techniques for Processing Multimedia Content, 1(2004): 9-16*.

Hadi, S., Fitrianah, D., Ayumi, V., & Lim, S. M. (2023). The Data Analysis of Determining Potential Flood-Prone Areas in DKI Jakarta Using Classification Model Approach. *International Journal of Application on Sciences, Technology and Engineering, 1(1): 347-357*.

Hall, M. J., Minns, A. W., & Ashrafuzzaman, A. K. M. (2002). The application of data mining techniques for the regionalisation of hydrological variables. *Hydrology and Earth System Sciences, 6(4): 685-694*.

Herman, E., Zsido, K. E., & Fenyves, V. (2022). Cluster analysis with k-mean versus k-medoid in financial performance evaluation. *Applied Sciences, 12(16): 7985*.

Istiadi, Y., & Priatna, D. (2021). Analysis of the determinants and typology of hydrometeorological disaster in Sukajaya Subdistrict, Bogor Regency, West Java, Indonesia. *Indonesian Journal of Applied Environmental Studies, 2(1): 41-46*.

Kaur, N. K., Kaur, U., & Singh, D. (2014). K-Medoid clustering algorithm-a review. *Int. J. Comput. Appl. Technol, 1(1): 42-45*.

Lackovic, M., Talia, D., & Palopoli, L. (2014). *Service-Oriented workflows for distributed data mining applications* (Doctoral dissertation).

Liu, C. R., Gibbs, C., & Coady, Y. (2005). SONAR: Customizable, Lightweight Tool Support to Prevent Drowning in Diagnostics. In *RAM-SE'05, 2nd ECOOP Workshop on Reflection, AOP and Meta-Data for Software Evolution*.

Lyons, Sthepanie. (2015). The Jakarta floods of early 2014: Rising risks in one of the World's fastest sinking cities. *Liège Université*, Viewed 15 October 2018.

Maghakian, C., Navratil, O., Zanot, J. M., Riviere, N., & Honegger, A. (2023). Drowning incidents in urban rivers: An underestimated issue with future challenges in need of an interdisciplinary database to characterise its epidemiology. Environmental Challenges.14. 100822. 10.1016/j.envc.2023.100822.

Napier, J. (2021). Living with Water: Infrastructure and Urbanism in Jakarta. *Ecocycles, 7(1): 52-72*.

Padawangi, R., & Douglass, M. (2015). Water, water everywhere: Toward participatory solutions to chronic urban flooding in Jakarta. *Pacific Affairs, 88(3): 517-550*.

Putri, D. P. (2018). UTILIZING DATA MINING IN GETTING INFORMATION FOR VEHICULAR USERS IN INDONESIA. *International Research Journal of Computer Science (IRJCS), 11(5): 525-535*.

Rafindadi, A. D., Shafiq, N., Othman, I., Ibrahim, A., Aliyu, M. M., Mikic, M., & Alarifi, H. (2023). Data mining of the essential causes of different types of fatal construction accidents. *Heliyon, 9(2)*.

Salim, W., Bettinger, K., & Fisher, M. (2019). Maladaptation on the waterfront: Jakarta's growth coalition and the Great Garuda. *Environment and Urbanization Asia, 10(1): 63-80*.

Shetty, P., & Singh, S. (2021). Hierarchical clustering: a survey. *International Journal of Applied Research, 7(4): 178-181*.

Sinaga, K. P., & Yang, M. S. (2020). Unsupervised K-means clustering algorithm. *IEEE Access, 8: 80716-80727*.

Sudirman, A. P., Windarto, & Wanto, A. (2018). Data Mining Tools | RapidMiner : K-Means Method on Clustering of Rice Crops by Province as Efforts to Stabilize Food Crops In Indonesia. IOP Conference Series: *Materials Science and Engineering, 420(12089): 1-8*.

Supriyadi, B., Windarto, A. P., & Soemartono, T. (2018). Classification of natural disaster prone areas in Indonesia using K-means. *International Journal of Grid and Distributed Computing, 11(8): 87-98*.

Utami, W., Wibowo, Y. A., Hadi, A. H., & Permadi, F. B. (2021). The impact of mangrove damage on tidal flooding in the sub district of Tugu, Semarang, Central Java. *Journal of Degraded and Mining Lands Management, 9(1): 3093*.

Ward, P. J., Budiyono, Y., & Marfai, M. A.: Flood risk in Jakarta, in: Severe Weather in Eastern Asia, Perils, Risks, Insurance. Munich Re Knowledge Series Natural Hazards, 106–107.

Wijayanti, P., Zhu, X., Hellegers, P., Budiyono, Y., & van Ierland, E. C. (2017). Estimation of river flood damages in Jakarta, Indonesia. *Natural hazards, 86: 1059-1079*.