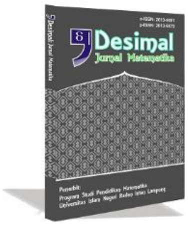
Desimal: Jurnal Matematika Vol 6 No 2 (2023) 261-272

Contents lists available at DJM
DESIMAL: JURNAL MATEMATIKA
p-ISSN: 2613-9073 (print), e-ISSN: 2613-9081 (online), DOI 10.24042/djm
http://ejournal.radenintan.ac.id/index.php/desimal/index



Clustering flood prone areas in deli serdang regency using density-based spatial clustering of applications with noise (dbscan) method

Nurmadani*, Fibri Rakhmawati

Universitas Islam Negri Sumatera Utara, Indonesia

ARTICLE INFO

Article History
Received : 11-07-2023
Revised : 10-08-2023
Accepted : 14-08-2023
Published : 30-08-2023

Keywords:
DBSCAN; Cluster; Silhouette Coefficient; Flood Prone Area Level.

*Correspondence: E-mail:
nurmadani0703192033@uinsu.ac.id

Doi:
10.24042/djm.v6i2.19043

ABSTRACT

Deli Serdang Regency is the most frequently flooded area in North Sumatra Province, causing many casualties and other losses to residents in flooded areas. Deli Serdang Regency has 22 sub-districts, each of which has a different level of flood vulnerability. Efforts are needed to categorize the level of flood vulnerability that needs to be watched out for in Deli Serdang Regency. The clustering used in this research is Density-Based Spatial Clustering Applications with Noise (DBSCAN). The purpose of this study is to determine the level of proneness to flooding in each region in 2022 in Deli Serdang Regency. The clustering results in this study concluded that using the DBSCAN algorithm, we obtained 2 clusters and 4 noise with a silhouette coefficient value of 0.395050089 with Epsilon 1.19 and MinPts 3. From the silhouette coefficient results, it can be concluded that the cluster structure obtained is weak because, with more variables, the calculation of distance based on density becomes invalid.

http://ejournal.radenintan.ac.id/index.php/desimal/index

INTRODUCTION

Flooding is defined as a condition where water in drainage channels (watersheds) cannot be absorbed or the flow of water in drainage channels is blocked so that it floods the surrounding floodplains. Flooding is an event that threatens and disrupts people's lives and livelihoods, resulting in damage to agricultural land, casualties, property losses, and psychological impacts (Aprillya & Chasanah, 2021). According to Indonesian Law No. 24 of 2007 (UU No. 24, 2007), preparedness is a series of activities carried out to anticipate disasters through organization and through appropriate, effective steps, and preparedness is actions that enable governments, organizations, communities, communications, and individuals to be able to respond to a disaster situation quickly and appropriately (Harahap et al., 2015).

Generally, flooding is caused by high rainfall intensity, as a result of which water drainage systems such as rivers, tributaries, drainage channel systems, and floodplain canals are unable to cope with the amount of accumulated rainwater so that the water overflows. The ability and capacity of this water drainage system are not always the same; this can occur due to changes such as sedimentation in the river, narrowing of the river due to human actions, blockage of garbage, and many other factors. The factors that can cause flooding are due to a lack of knowledge and readiness in disaster management. So knowledge of flood disaster mitigation and prevention is needed to increase public knowledge and provide information on disaster management in order to reduce the impact of damage that can be caused by floods (Septian et al., 2020). Several other factors that cause flooding include rainfall over a long period of time, soil erosion that leaves rocks and no water infiltration, blockage of water flow due to poor handling of garbage and instead throwing it into the water, damaged dams and waterways, illegal and uncontrolled logging, the topology of an area of shipment or due to flash floods, and land conversion into settlements and offices (Tampubolon, 2018). In other studies, factors that cause flooding include rainfall, elevation (screen height), population density, and distance from the river (Sukmayadi et al., 2021). Floods can have an impact on social life, the economy, material damage, environmental damage, and cultural heritage (Setiawan et al., 2020).

A country with a tropical climate ensures that Indonesia has rainfall and humid temperatures. This poses a threat to hydrometeorological disasters. North Sumatra is an area prone to flooding. Floods occur due to the inundation of water in a certain place within a certain period of time (I. H. O. Sitorus et al., 2021). Indonesia is also one of the countries with a high vulnerability to various disasters, especially floods. In the tropics, these two phenomena usually cause shifts in rainfall patterns and temperature changes that result in long dry seasons or prolonged rainy seasons that can cause flooding in various places. North Sumatra has high rainfall intensity and varies throughout the year (Tampubolon, 2018). In addition to rainfall, flooding is also caused by other factors such as regional topography, geological structure, water bodies (drainage), and environmental changes (Tasia & Afdal, 2023).

In 2021, there are several regencies and cities in North Sumatra that have the potential to be flooded. Flood areas based on flood-prone maps in North Sumatra include Medan City, Tebing Tinggi City, Tanjung Balai City, Tanjung Pura City, Sei Rampah City, Pematang Siantar City, Natal City, Asahan Regency, Serdang Bedagai Regency, and Langkat Regency. Tebing Tinggi City is one of the areas in North Sumatra province that has the potential for flood disasters to occur. Based on BMKG data, the Tebing Tinggi City area (Kec. Bajenis, Tebingtinggi City, Padang Hilir, Padang Hulu, Rambutan) has a medium level of potential flood-prone category (Sitorus, 2022). Regions in North Sumatra generally have a fairly high level of vulnerability to flood disasters (Damanik & Restu, 2012).

Until now, Sumatra province has been one of the most vulnerable areas prone to flooding, including in Deli Serdang district, due to high rainfall, humid air temperatures, and overflowing water where the capacity of the river cannot accommodate the amount of water. Therefore, this problem will be solved by grouping flood-prone areas, namely areas that are often flooded, with areas that are rarely or never flooded. This clustering can be used with the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) method. This method is a clustering method that categorizes the high density of a cluster area found in a free form by utilizing noise (Hatta et al., 2021).

Clustering is the most common type of supervised learning element used in data mining. Clustering is done to organize data into a finite set of systematically consistent groups based on similarity calculations. The clustering process aims to group data that has the same characteristics in the same cluster and data that has different characteristics is grouped into other clusters (Rahman & Wijayanto, 2021). Clustering of flood-prone areas is very necessary as input to the government so that it can take targeted policies in an effort to overcome it. Flood management has been carried out in various cities in the world, including Indonesia, such as through drainage, infiltration wells, and various other methods, but they do not function optimally due to human activities (Anggraini et al., 2021).

In data mining, there are many popular algorithms, one of which is density-based spatial clustering of applications with noise (DBSCAN). DBSCAN is an algorithm that is included in the density-based clustering category; namely, the cluster formation process is carried out based on the level of proximity or distance density between objects in the dataset (Rahman & Wijayanto, 2021). Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is one of the clustering methods that can find outlier data in a data set. DBSCAN works by determining clusters based on data density using the epsilon (range) and MinPts (minimum point to form a cluster) parameters (Furqon & Muflikhah, 2016). The method that will be used in the clustering is DBSCAN, which is analyzed using the Silhouette Coefficient, epsilon value, and MinPts (Furqon & Muflikhah, 2016). The visualization used in the implementation of the data mining process using the DBSCAN algorithm is RStudio (Hermanto & Sunandar, 2020). This research uses the DBSCAN algorithm with several considerations, including its ability to detect outliers and noise, the fact that it does not need to know the number of clusters (k) that will be formed, and its ability to recognize irregular cluster shapes.

In addition, it refers to several previous studies by Jatipaningrum, Azhari, & Suryowati (2022) that concluded that the best method is the DBSCAN method compared to K-Means using Manhattan distance with MinPts = 2 and epsilon = 4, which has the smallest DBI value of 0.284 and produces less noise (Tasia & Afdal, 2023). The K-Means method is more optimal than the K-Medoids method. Kristianto (2021) concluded that the Silhouette Coefficient results show that DBSCAN has better performance than K-Means clustering with value of 0.99 (Schubert et al., 2017a). It was concluded that the original DBSCAN with an effective index and reasonably selected parameter values worked competitively compared to the method proposed by Gan and Tao (Schubert et al., 2017b). Isnarwaty & Irhamah (2020) concluded that based on the silhouette coefficient value, the DBSCAN method is better than K-Means in clustering tweets shown to JNE, J&T, and Pos Indonesia expedition services because it produces a higher silhouette coefficient. Showing that DBSCAN is able to form clusters well with an effective index and reasonable parameter values, it works competitively against other methods that are 10 to 1,000 times more runtime (Rahman & Wijayanto, 2021). The purpose of this research is to cluster flood-prone areas in Deli Serdang Regency based on the characteristics of the resulting flood-prone area parameters so that one day it can be used as recommendation material for flood disaster mitigation for related institutions (Rahman & Wijayanto, 2021). The purpose of this research is to group flood-prone areas in Deli Serdang Regency based on the characteristics of the resulting flood-prone area parameters.

From the description above and from some of these studies, the author proposes a study entitled "Grouping Flood-Prone Areas in Deli Serdang Regency Using the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) Method". With this research, it is hoped that it can provide a better solution and be a reference for policies that will be taken by BMKG in decision-making.

METHOD

Dataset

The area studied is Deli Serdang district, which consists of 22 sub-districts. The visualization used in the implementation of the data mining process using the DBSCAN algorithm is RStudio. The methods used are clustering, DBSCAN, and Silhouette Coefficients. The research period was taken in 2022. This research uses four variables, namely elevation and slope obtained from Shuttle Radar Topography Mission (SRTM) data processing, rainfall obtained from BMKG Sampali Medan, land use parameters obtained from Sumatra Province Water Resources (SDA), and population density obtained from the BPS Deli Serdang website (Badan Pusat Statistika, 2023).

The purpose of this research, based on the formulation of the problem, is to determine the results of the cluster formed by using the density-based spatial clustering method of application with noise (DBSCAN) for grouping flood-prone areas in Deli Serdang district.

Data Preprocessing

The initial data on flood-prone areas has complete data and some is incomplete, including data that has double data and empty data. In the data mining process, such data will give good results (Hermanto & Sunandar, 2020). Data mining is often called knowledge discovery in databases (KDD), which means activities that include collecting and using historical data to find regularities, patterns, or relationships in large datasets (Rohalidyawati et al., 2020). KDD is related to the visualization of patterns in a number of data sets used in the implementation of the data mining process using the DBSCAN algorithm using the R-Studio program (Hermanto & Sunandar, 2020). Data mining has the aim of analyzing and processing data into information using several standard techniques, namely data classification, data clustering, ontology and taxonomy creation, document summarization, and latent corpus analysis (Hatta et al., 2021).

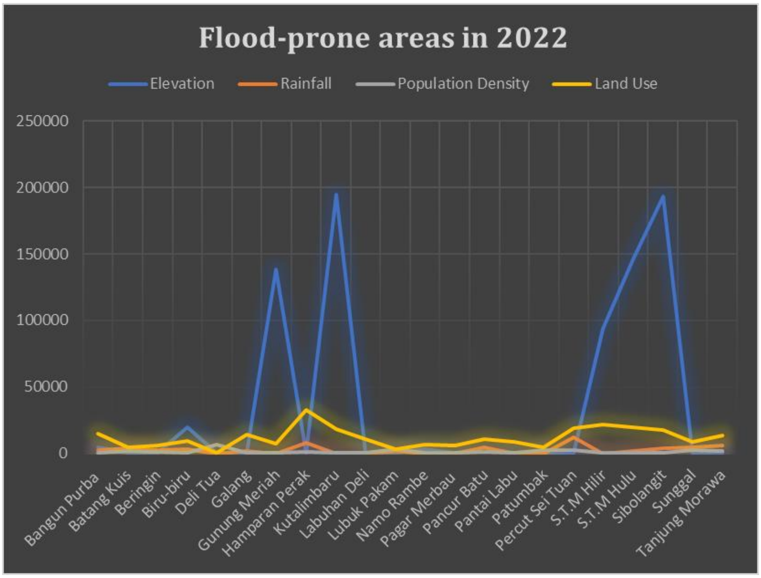


Figure 1. Variable Graph of Flood-Prone Areas in 2022

Preprocessing is needed to ensure the data is appropriate in order to produce the right analysis. The distance calculation stage between data points in the DBSCAN method uses the Euclidean Distance function, so that the difference in the range of values between variables greatly affects the clustering results. Therefore, in this study, a normalization process was carried out to equalize the range of values of all variables using the Z-Score normalization function with the following equation (Rahman & Wijayanto, 2021):

$$\text{Zscore} = \frac{x - \mu_x}{\sigma_x} \quad (1)$$

where score is the normalized result, x is the i-th data, μ_x is the average value for variable x, and σ_x is the standard deviation of variable x.

Clustering

Clustering is a data mining technique that groups similar objects (Mahendra et al., 2021). The distance calculation stage between data points in the DBSCAN algorithm uses the Euclidean distance function so that the difference in the range of values between variables greatly affects the clustering results. The following is the Euclidean distance formula shown in Equation (2) (Jatipaningrum et al., 2022):

$$d_{ij} = \sqrt{\sum_{k=1}^{p} (x_{ik} - y_{jk})^2} \quad (2)$$

where d_{ij} is the distance between the i-th object and the center of the *cluster* j, P is the number of cluster variables, x_{ik} is the data from the object to -i on the k-th variable, and $y_{jK}$is the center of the j-th cluster on the k -th variable.

DBSCAN Algorithm

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is an algorithm to cluster data based on data density.

There are several components or terms contained in the DBSCAN algorithm (Harjanto et al., 2021). Epsilon: is the number of points in a certain radius

1) Noise Point: the outermost point of the density or (Epsilon).
2) Border Point: border points have less than MinPts in Epsilon, but are still within the neighborhood of the core point.
3) MinPts number of nearest neighbors is used to define the local neighborhood of an object.
4) Core Point: A point in the interior of the cluster.

The concept of density referred to in DBSCAN is the minimum amount of data within a radius epsilon, the data belongs to the desired density category.

The algorithm steps of DBSCAN are as follows (Wijaya et al., 2021):

1. Determining the MinPts and Epsilon parameters on the scatter data.
2. Determine the value of p, or starting point, randomly.
3. Count the number of points specified by the radius parameter (epsilon). If the number is sufficient (greater than or equal to ε), the data will be marked as core points. If the points are within the radius $\varepsilon \leq$ MinPts, then noise occurs. but if $\varepsilon \geq$ MinPts, then a cluster is formed.
4. Calculate the distance between core points and other points using Euclidean distance, which can be used by calculating the closest distance from data to a centroid point. The Euclidean distance calculation can be used to calculate the similarity level of samples.
5. Create a new cluster by adding points p to the cluster, i.e., identifying the data marked as core points. Then continue the process until all points have been processed. If there are points that do not fit into any cluster, they will be marked as noise.
6. Repeat the steps iteratively for other data distributions.

The advantages of the DBSCAN method include (Nurhaliza & Mustakim, 2021):

a. DBSCAN does not require information about the number of groups to be formed.
b. DBSCAN can find arbitrarily shaped clusters and can even find clusters that are completely surrounded by (but not connected to) different clusters.
c. DBSCAN has information about the noise.
d. DBSCAN requires only two parameters that are largely insensitive to the order of points in the database.

Cluster Validation

To avoid discrepancies in results and ensure that cluster results reflect the general population, clustering results must be validated. One method to test the validation of cluster results is to use the Silhouette Coefficient. Silhouette Coefficient validation formula (Harjanto et al., 2021):

$$s(i) = \frac{b(i)-a(i)}{\max\{a(i)b(i)\}} \quad (3)$$

where $a(i)$ is the average distance of point i to all objects in cluster $a(i)$ and $b(i)$ is the average distance of point i to all objects in other clusters.

The silhouette coefficient is obtained from the largest value of each silhouette value and whether or not the Silhouette Coefficient value is valid as measured in Table 1, which shows the structure of a cluster.

Table 1. Silhouette Coefficient Structure

Silhouette Coefficient Value	Interpretation of Silhouette Coefficients
$0.7 < SC \leq 1$	Strong Structure
$0.5 < SC \leq 0.7$	Good Structure
$0.25 < SC \leq 0.5$	Weak Structure
$SC \leq 0.25$	Poor Structure

The process of this analysis method is explained in the flowchart, as shown in Figure 2.

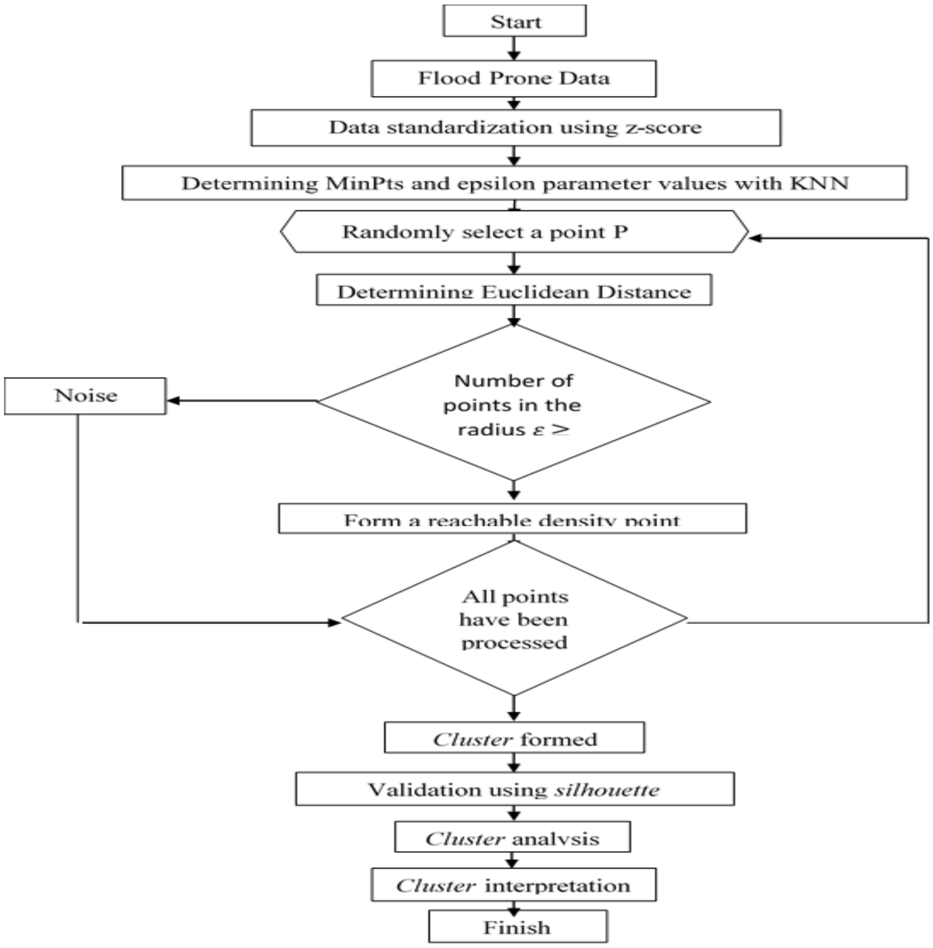


Figure 2. DBSCAN Flowchart

RESULTS AND DISCUSSION

Determining epsilon and MinPts

After the data is standardized using the z-score, determine the values of ε and MinPts. DBSCAN requires two input parameters to be able to cluster, namely Epsilon for the radius that determines the neighborhood boundary of the point (Epsilon-neighborhood) and MinPts (Mahendra et al., 2021). The KNN algorithm is a method used to classify data based on the shortest distance to data objects, determining the best k value for KNN based on existing data (Cholil et al., 2021). The result of the knee point used as epsilon, KNN, in determining epsilon is also called a *k-dist graph* because the process requires graph observation. The values ε used in the analysis are determined using the *k-Nearest Neighbor* algorithm with $k = 3$.

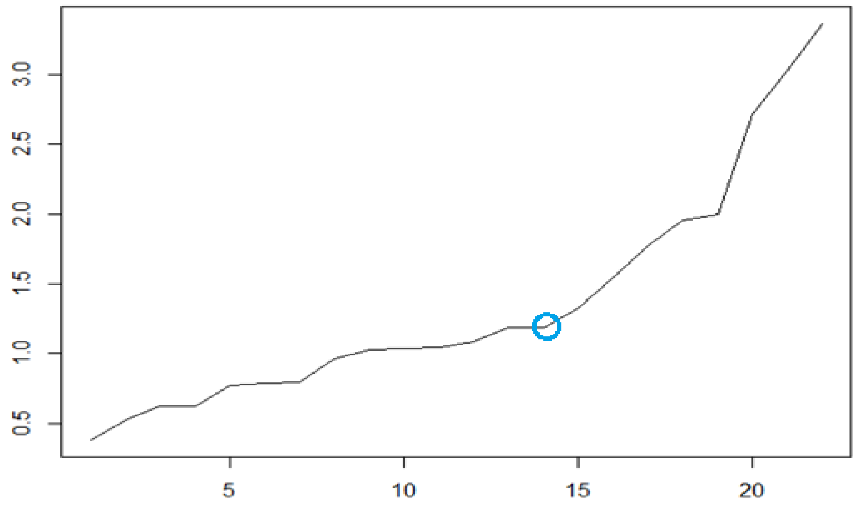


Figure 3. Plot of 3-NN on Determination of $\boldsymbol{\varepsilon}$

The KNN plot shows that the optimal epsilon value is in the range of 1.19 points through the observation of the knee of a curve.

The knee of a curve is the point where the curve bends, especially from a high slope to a low slope or in the other direction. Used in optimization if the points are optimal for some decision, for example, when there is an increase from vertical to horizontal.

Table 2. Combination of Input Values $\boldsymbol{\varepsilon}$ and MinPts

Epsilon	MinPts	Cluster	Noise	Silhouette
0.98	3	2	5	0.34
1.19	**3**	**2**	**4**	**0.40**
2.45	3	1	1	0.45
2.66	3	1	1	0.45
2.70	3	1	1	0.45

Based on the results of several experiments, an epsilon value of 1.19 and MinPts 3 with a Silhouette value of 0.40 were selected. In some of these experiments, we did not take the epsilon with the highest Silhouette value because there was only one cluster and the location of the elbow point on MinPts 3 was located at an epsilon value of 1.19. The purpose of cluster analysis is to group objects based on their characteristics.

Clustering using the DBSCAN algorithm

After obtaining the optimum values of ε and MinPts optimum, which is characterized by a Silhouette Coefficient value of 0.395050089. Next, the clustering process is carried out by randomly selecting p to form a cluster using the DBSCAN algorithm. If all points have been processed, then the cluster is formed. The values of ε and MinPts (ε=1.19 and MinPts=3) are used as input to run the clustering process in DBSCAN. The clustering contains 2 cluster (s) and 4 noise points.

Table 3. DBSCAN Output using Minpts = 3 and $\boldsymbol{\varepsilon}$ = 1.19

Output	Description	Total
0	Noise	4
1	Cluster 1	14
2	Cluster 2	4

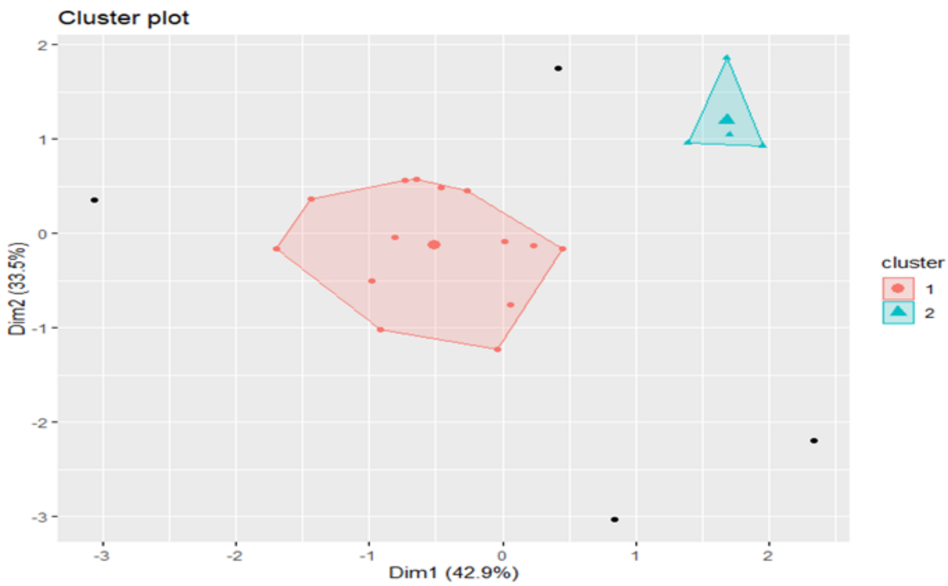


Figure 4. DBSCAN Cluster Plot

It can be seen that there are 2 clusters produced: the first cluster has 14 points, the second cluster has 4 points, and noise has 4 points. Red is a point from cluster 1, blue is a point from cluster 2, and black points are noise. Red points are density connected, indicating that cluster 1 is formed; red blue points are density connected, indicating that cluster 2 is formed; and black points have no density connected points, which is noise.

The Silhouette Coefficient calculation uses data that has formed clusters with equation 2 on all data. Then look for the Euclidean distance to find the value of $a(i)$ and $b(i)$, where $a(i)$ is the average distance of the same cluster and $b(i)$ is the average distance in other clusters. then the results of the silhouette value are shown in Table 4.

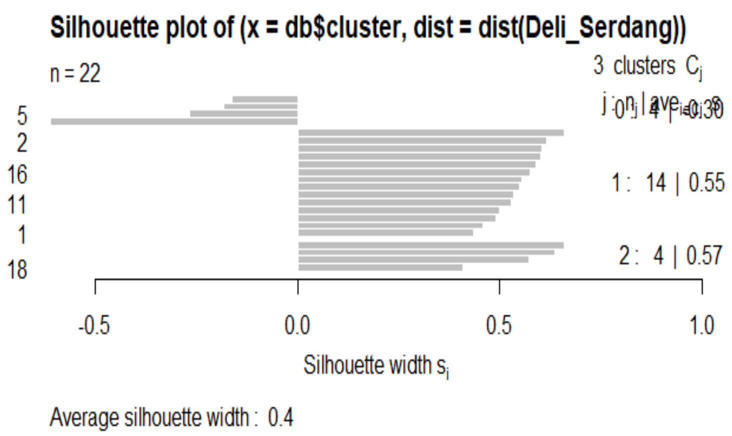
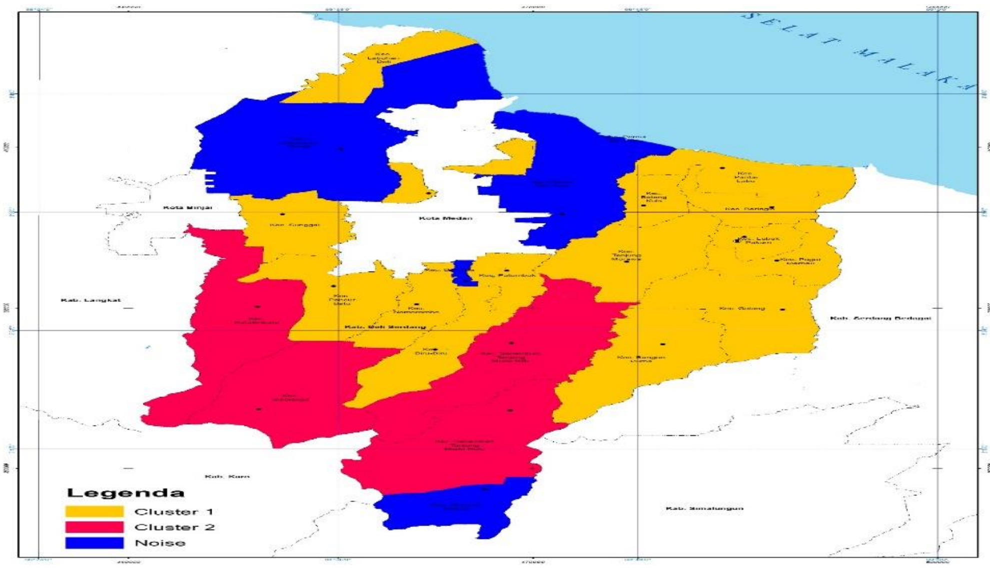


Figure 5. Silhouette Coefficient Values in the R Program

Table 4. Silhouette Count

Cluster	Subdistrict	$a(i)$	$b(i)$	$s(i)$
Cluster 1	Bangun Purba	1,383563143	2,454845833	0,436395099
	Batang Kuis	1,320674013	3,394282201	0,610912135
	Beringin	1,066730244	3,140618735	0,660343921
	Biru-Biru	1,158108976	2,59502862	0,553720153
	Galang	1,266225185	2,523239774	0,498174847
	Labuhan Deli	1,230212949	2,72666149	0,5488208
	Lubuk Pakam	1,682302524	3,505283071	0,520066571
	Namo Rambe	1,183426635	2,964108928	0,600747927
	Pagar Merbau	1,218071627	3,063135081	0,602344789
	Pancur Batu	1,353461748	2,907393282	0,534475863
	Pantai Labu	1,168944284	2,850052724	0,589851698
	Patumbak	1,46421856	3,445889684	0,575082578
	Sunggal	1,645999847	3,176750793	0,48186057
	Tanjung Morawa	1,658949898	3,029234581	0,45235344
Cluster 2	Kutalimbaru	1,23792535	3,411015108	0,637080074
	S.T.M Hilir	1,501200354	2,53333433	0,407421146
	S.T.M Hulu	0,974272012	2,848202583	0,65793444
	Sibolangit	1,434473356	3,344564591	0,571103109
Noise	Deli Tua	5,569804869	3,909544365	-0,298082346
	Gunung Meriah	4830291846	1,843838973	-0,618275866
	Hamparan Perak	4496838701	3,702808284	-0,17657525
	Percut Sei Tuan	4239710863	3,584023697	-0,154653746
	Silhoutte Coefficient			**0,395050089**

Figure 5 has data with a negative value, namely in the noise data due to the nature of the noise, which is too far away from the other data. This Silhouette Coefficient result shows a weak structure because there are four results that cause a negative Silhouette Coefficient value.



Description: yellow = cluster1 pink=cluster2 blue=cluster3

Figure 6. Cluster Map of Flood-Prone Areas in 2022

Cluster Interpretation

The initial step using the DBSCAN method is to determine the value of epsilon = *1*.19 and MinPts = 3 obtained from *k-distance*. After the epsilon and MinPts values are obtained, proceed to random centroid selection. After knowing the initial centroid value, the distance between sub-districts is calculated using Euclidean. From the results of this distance, if a point is selected that is less than epsilon and more than MinPts, then it will form a cluster, but if it is more than epsilon and less than MinPts, then the point becomes noise. Processing continues until all points are processed. Here are the final results of cluster formation:

Table 5. Final Centroid Table of DBSCAN Cluster Formation

District	Year 2022
Bangun Purba Sub-district	Low
Biru-biru Sub-district	Low
Galang Subdistrict	Low
Labuhan Deli Sub-district	Low
Pancur Batu Sub-district	Low
Pantai Labu Sub-district	Low
Batang Kuis Sub-district	Low
Beringin Sub-district	Low
Lubuk Pakam Sub-district	Medium
Namo Rambe Sub-district	Medium
Pagar Merbau Sub-district	Medium
Patumbak Sub-district	Medium
Sunggal Sub-district	Low
Tanjung Morawa Sub-district	Low
Lower S.T.M District	Medium
S.T.M Hulu Sub-district	Medium
Kutalinbaru Subdistrict	High
Sibolangit Sub-district	High
Deli Tua Sub-district	High
Gunung Meriah Sub-district	Medium
Sub-district Hamparan Perak	High
Percut Sei Tuan Sub-district	High

In Table 5, the sub-districts are included in the low, medium, and high flood-prone areas in 2022. It can be seen that Deli Tua Subdistrict, Hamparan Perak Subdistrict, and Percut Sei Tuan Subdistrict are always the noise, as well as subdistricts with a high level of frequent flooding. This means that these three sub-districts need the government's attention to overcome the most flood-prone areas in Deli Serdang Regency.

CONCLUSIONS AND SUGGESTIONS

Based on the level of flood-prone areas in Deli Serdang Regency using the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) method, the Silhouette Coefficient value in 2022 is 0.40, which is included in the weak structure with a range of 0.26-0.50 in Table 1. The Silhouette Coefficient validation is obtained with a weak structure because, with more variables, the calculation of distance based on density becomes invalid. From the results of the cluster level of flood-prone areas, 2 clusters with noise were obtained, namely in Deli Tua District, Gunung Meriah District, Hamparan Perak District, and Percut Sei Tuan District in 2022. The district that becomes the noise is the one that has the highest level of flood-prone areas in Deli Serdang Regency.

Suggestions for future researchers Research is needed on the development of the DBSCAN method, one of which is the HDBSCAN method, and this method should use spatial-based data.

REFERENCES

Anggraini, N., Pangaribuan, B., Siregar, A. P., Sintampalam, G., Muhammad, A., Damanik, M. R. S., & Rahmadi, M. T. (2021). Analisis pemetaan daerah rawan banjir di kota medan tahun 2020. *Jurnal Samudra Geografi*, *4*(2), 27–33. https://doi.org/10.33059/jsg.v4i2.3851

Aprillya, M. R., & Chasanah, U. (2021). Analisis lahan pertanian rawan banjir menggunakan metode multi atribut utility theory berbasis sistem informasi geografis. *Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer*, *16*(2). https://doi.org/10.30872/jim.v16i2.6554

Badan Pusat Statistika. (2023). *Kabupaten deli serdang dalam angka*. https://deliserdangkab.bps.go.id/publication.html?page=3

Cholil, S. R., Handayani, T., Prathivi, R., & Ardianita, T. (2021). Implementasi algoritma klasifikasi k-nearest neighbor (knn) untuk klasifikasi seleksi penerima beasiswa. *IJCIT (Indonesian Journal on Computer and Information Technology)*, *6*(2), 118–127.

Damanik, M. R. S., & Restu, R. (2012). Pemetaan tingkat risiko banjir dan longsor sumatera utara berbasis sistem informasi geografis. *Jurnal Geografi*, *4*(1), 29–42.

Furqon, M. T., & Muflikhah, L. (2016). Clustering the potential risk of tsunami using density-based spatial clustering of application with noise (dbscan). *Journal of Enviromental Engineering and Sustainable Technology*, *3*(1), 1–8. https://doi.org/10.21776/ub.jeest.2016.003.01.1

Harahap, M. E., Lufti, M., & Muthalib, A. (2015). Pengaruh pengetahuan dan sikap terhadap kesiapsiagaan masyarakat dalam menghadapi banjir di desa perkebunan bukit lawang kecamatan bahorok tahun 2011. *Jurnal Ilmiah Keperawatan Imelda*, *1*(1), 22–31.

Harjanto, T. D., Vatresia, A., & Faurina, R. (2021). Analisis penetapan skala prioritas penanganan balita stunting menggunakan metode dbscan clustering (studi kasus data dinas kesehatan kabupaten lebong). *Rekursif: Jurnal Informatika*, *9*(1). https://doi.org/10.33369/rekursif.v9i1.14982

Hatta, M. S., Azmi, F., & Setianingsih, C. (2021). Clustering pada data sentimen penggunaan transportasi online menggunakan algoritma spectral clustering. *E-Proceedings of Engineering*, *8*(6).

Hermanto, T. I., & Sunandar, M. A. (2020). Analisis data sebaran penyakit menggunakan algoritma density based spatial clustering of applications with noise. *Jurnal Sains Komputer Dan Teknologi Informasi*, *3*(1). https://doi.org/10.33084/jsakti.v3i1.1775

Isnarwaty, D. P., & Irhamah, I. (2020). Text clustering pada akun twitter layanan ekspedisi jne, j&t, dan pos indonesia menggunakan metode density-based spatial clustering of applications with noise (dbscan) dan k-means. *Jurnal Sains Dan Seni ITS*, *8*(2). https://doi.org/10.12962/j23373520.v8i2.49094

Jatipaningrum, M. T., Azhari, S. E., & Suryowati, K. (2022). Pengelompokan kabupaten dan kota di provinsi jawa timur berdasarkan tingkat kesejahteraan dengan metode k-means dan density-based spatial clustering of applications with noise. *Jurnal Derivat: Jurnal Matematika Dan Pendidikan Matematika*, *9*(1), 70–81. https://doi.org/10.31316/j.derivat.v9i1.2832

Kristianto, A. (2021). Analisa performa k-means dan dbscan dalam clustering minat penggunaan transportasi umum. *Elkom : Jurnal Elektronika Dan Komputer*, *14*(2), 368–372. https://doi.org/10.51903/elkom.v14i2.551

Mahendra, R., Azmi, F., & Setianingsih, C. (2021). Klasterisasi pada data penggunaan listrik di gedung telkom university menggunakan algoritma density-based spatial clustering of application with noise (dbscan). *E-Proceedings of Engineering*, *8*(6).

Nurhaliza, N., & Mustakim, M. (2021). Pengelompokan data kasus covid-19 di dunia menggunakan algoritma dbscan. *IJIRSE: Indonesian Journal of Informatic Research and Software Engineering*, *1*(1), 1–8.

Rahman, R. R. A., & Wijayanto, A. W. (2021). Pengelompokan data gempa bumi menggunakan algoritma dbscan. *Jurnal Meteorologi Dan Geofisika*, *22*(1).

https://doi.org/10.31172/jmg.v22i1.738

Rohalidyawati, W., Rahmawati, R., & Mustafid, M. (2020). Segmentasi pelanggan e-money dengan menggunakan algoritma dbscan (density based spatial clustering applications with noise) di provinsi dki jakarta. *Jurnal Gaussian*, *9*(2), 162–169. https://doi.org/10.14710/j.gauss.v9i2.27818

Schubert, E., Sander, J., Ester, M., Kriegel, H. P., & Xu, X. (2017a). Dbscan revisited, revisited: Why and how you should (still) use dbscan. *ACM Transactions on Database Systems*, *42*(3), 1–21. https://doi.org/10.1145/3068335

Schubert, E., Sander, J., Ester, M., Kriegel, H. P., & Xu, X. (2017b). Why and How You Should (Still) Use DBSCAN. *ACM Transactions on Database Systems*, *42*(3), 1–21.

Septian, A., Elvarani, A. Y., Putri, A. S., Maulia, I., Damayanti, L., Pahlevi, M. Z., & Aswad, F. H. (2020). Identifikasi zona potensi banjir berbasis sistem informasi geografis menggunakan metode overlay dengan scoring di kabupaten agam, sumatera barat. *Jurnal Geosains Dan Remote Sensing*, *1*(1). https://doi.org/10.23960/jgrs.2020.v1i1.25

Setiawan, Y., Purwandari, E. P., Wijanarko, A., & Sunandi, E. (2020). Pemetaan zonasi rawan banjir dengan analisis indeks rawan banjir menggunakan metode fuzzy simple adaptive weighting. *Pseudocode*, *7*(1), 78–87. https://doi.org/10.33369/pseudocode.7.1.78-87

Sitorus, I. H. O., Bioresita, F., & Hayati, N. (2021). Analisa tingkat rawan banjir di daerah kabupaten bandung menggunakan metode pembobotan dan scoring. *Jurnal Teknik ITS*, *10*(1). https://doi.org/10.12962/j23373539.v10i1.60082

Sitorus, S. B. A. (2022). *Peran badan penanggulangan bencana daerah dalam penanggulangan banjir di kota tebing tinggi provinsi sumatera utara*. Institut Pemerintahan Dalam Negeri.

Sukmayadi, C., Primajaya, A., & Maulana, I. (2021). Penerapan algoritma k-medoids dalam menentukan daerah rawan banjir di kabupaten karawang. *INFORMAL: Informatics Journal*, *6*(3), 187. https://doi.org/10.19184/isj.v6i3.25423

Tampubolon, K. (2018). *Aplikasi sistem informasi geografis (sig) sebagai penentuan kawasan rawan banjir di kota medan*.

Tasia, E., & Afdal, M. (2023). Perbandingan algoritma k-means dan k-medoids untuk clustering daerah rawan banjir di kabupaten rokan hilir. *Indonesian Journal of Informatic Research and Software Engineering (IJIRSE)*, *3*(1), 65–73. https://doi.org/10.57152/ijirse.v3i1.523

UU No. 24. (2007). *Undang-undang (uu) nomor 24 tahun 2007 tentang penanggulangan bencana*.

Wijaya, D. V., Alfiansyah, Y. F., Junior, A. R., Masruriyah, A. F. N., Indra, J., Hikmayanti, H., & Siregar, A. M. (2021). Analisis sentimen pada buletin menggunakan algoritme dbscan. *Conference on Innovation and Application of Science and Technology (CIASTECH)*.