Bulletin of Applied Mathematics and Mathematics Education
Vol. 6, No. 1, April 2026, pp. 69-84
e-ISSN: 2776-1029 p-ISSN: 2776-1002 DOI: 10.12928/bamme.v6ix.14910

APPLIED MATHEMATICS

Spatial clustering analysis of DKI Jakarta criminal cases in 2024 using DBSCAN and comparison with K-Means

Sintia Afriyani [1], Winarsi J. Bidul [1,*], Imam Al Maksur [1], Achmad Fauzan [1], Roza Azizah Primatika [2]

[1] Universitas Islam Indonesia, Jl. Kaliurang, Sleman, Daerah Istimewa Yogyakarta 55584, Indonesia
[2] Universitas Gajah Mada, Jl. Fauna, Karangmalang, Daerah Istimewa Yogyakarta 55281, Indonesia
*Corresponding author

Abstract

The high crime rate in DKI Jakarta requires spatial analysis to accurately identify vulnerable zones. Such information is essential for developing data-driven crime prevention strategies. Therefore, this study aims to map the spatial distribution of criminal cases in DKI Jakarta in 2024 and to evaluate two clustering methods Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and K-Means in order to determine the most effective approach for identifying crime-prone areas.

Data came from the Jakarta Open Data portal, containing coordinates (latitude, longitude), crime types, and supporting details. Pre-processing involved removing duplicates, filtering 2024 records, and eliminating invalid coordinates. Spatial features were normalized using Standard Scaler. DBSCAN parameters (eps, min_samples) were tuned via grid search, while K-Means used the Elbow method to determine optimal clusters. Performance was evaluated using Silhouette Score, Davies-Bouldin Index, and Calinski-Harabasz Index.

K-Means achieved a higher Silhouette Score (0.483), lower Davies-Bouldin Index (0.725), and higher Calinski-Harabasz Index (145.897), indicating more compact, well-separated clusters. DBSCAN formed more clusters (12) and identified noise points, showing its ability to capture spatial density variations and detect small-scale hotspots.

In conclusion, K-Means is more suitable for macro-level mapping of security areas, supporting the allocation of police resources and administrative planning, while DBSCAN is more effective in identifying localized hotspots that require targeted surveillance and rapid response. These findings provide practical insights for policymakers and law enforcement agencies in developing data-driven strategies for urban crime prevention and security management.

Keywords: spatial clustering, DBSCAN, K-Means, crime, DKI Jakarta

How to cite: Afriyani, S., Bidul, W. J., Al Maksur, I., Fauzan, A., & Primatika, R. A. (2026). Spatial clustering analysis of DKI Jakarta criminal cases in 2024 using DBSCAN and comparison with K-Means, *6*(1), 69-84. https://doi.org/10.12928/bamme.vxix.14910

Article history: Received 07/11/2025, Accepted 10/07/2026, Published 16/07/2026

Correspondence address: Universitas Islam Indonesia, Jl. Kaliurang, Sleman, Daerah Istimewa Yogyakarta 55584, Indonesia. E-mail: 24928005@students.uii.ac.id

INTRODUCTION

Crime is one of the social problems that continues to be a concern in urban areas, especially in metropolitan cities with high population density such as the Special Capital Region (DKI) Jakarta. The complexity of urban life, rapid population growth, and socio-economic disparities are factors that trigger the occurrence of various criminal acts, ranging from theft and robbery to acts of violence (Yusuf et al., 2025). Crimes not only threaten public safety, but also have an impact on the quality of life and social stability (Stiawan & Yusuf, 2025). Therefore, identifying spatial patterns of criminal incidence is an important step in formulating effective prevention strategies and optimizing the resources of law enforcement officials (Munajat & Yusuf, 2024).

Spatial analysis of crime data has received increasing attention in recent years, along with the development of the smart city concept and evidence-based *policy*) (Julian & Umar, 2025). The DKI Jakarta Provincial Government through the *Jakarta Open Data* portal provides public complaint data that can be used to map the location of criminal incidents (DKI Jakarta Provincial Government, 2023). With this mapping, security forces can identify hotspots or zones with high risk levels, and take more targeted preventive measures, such as increasing the frequency of patrols or adding security infrastructure (Muin et al., 2023).

In the realm of *data mining*, clustering is a method used to group data based on similar characteristics without requiring an initial label (Septianto et al., 2025). In research conducted by [7] For spatial data, the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm is used because it is able to detect clusters with irregular shapes and identify *outliers*, making it suitable for analyzing complex spatial distributions (Miftahurrahmi et al., 2024). However, DBSCAN requires precise parameter determination and can have difficulties if the data has varying densities (Salman, 2023). On the other hand, K-Means is a popular clustering method that is simpler and computationally efficient, but it assumes a spherical cluster shape and requires pre-determining the number of clusters, which may limit its performance on complex spatial data (Syahri et al., 2023).

Based on this background, this study aims to analyze the spatial clustering of criminal cases in DKI Jakarta in 2024 by applying the DBSCAN method and comparing it with K-Means. The novelty of this research lies in the integration of two clustering approaches to evaluate their performance on recent spatial crime data at the provincial level, which has not been widely explored in previous studies. This comparison is expected to provide a more comprehensive understanding of crime distribution patterns and serve as practical input for law enforcement officials, local governments, and other stakeholders in formulating data-driven crime prevention and control strategies.

RESEARCH METHOD

This study uses secondary data in the form of criminal complaint reports in the DKI Jakarta area during the January-December 2024 period to identify crime-prone zones. The data is sourced from the official Jakarta Open Data portal (https://data.jakarta.go.id) managed by the DKI Jakarta Provincial Government, which contains a compilation of reports from various channels such as complaint applications, emergency telephones, and direct reports to the authorities. The dataset contains information on the location of the incident in the form of address and geographical coordinates, type of crime, number of reports, and other supporting attributes.

The research variables consist of Id as the identification of the report, Number of complaints as the number of reports received for one incident, Period_data indicating the year of the incident, Region as the administrative area, Location_complaint in the form of incident address, Latitude and Longitude for geographic coordinates, Date_complaint as the time of the report, Origin_complaint representing the reporting channel, and Type_criminal as the category of crime. Data pre-processing and clustering analysis were carried out using Python programming language (Google Colab) with supporting libraries such as Pandas, Scikit-learn, Seaborn, and Folium for data processing, clustering, visualization, and interactive mapping. The analysis is conducted spatially using DBSCAN and K-Means clustering methods, with stages ranging from data pre-processing, normalization, standardization, optimal parameter determination, to model evaluation using Silhouette Score, Davies-Bouldin Index, and Calinski-Harabasz Index metrics. The research flow is depicted in Figure 1 below:

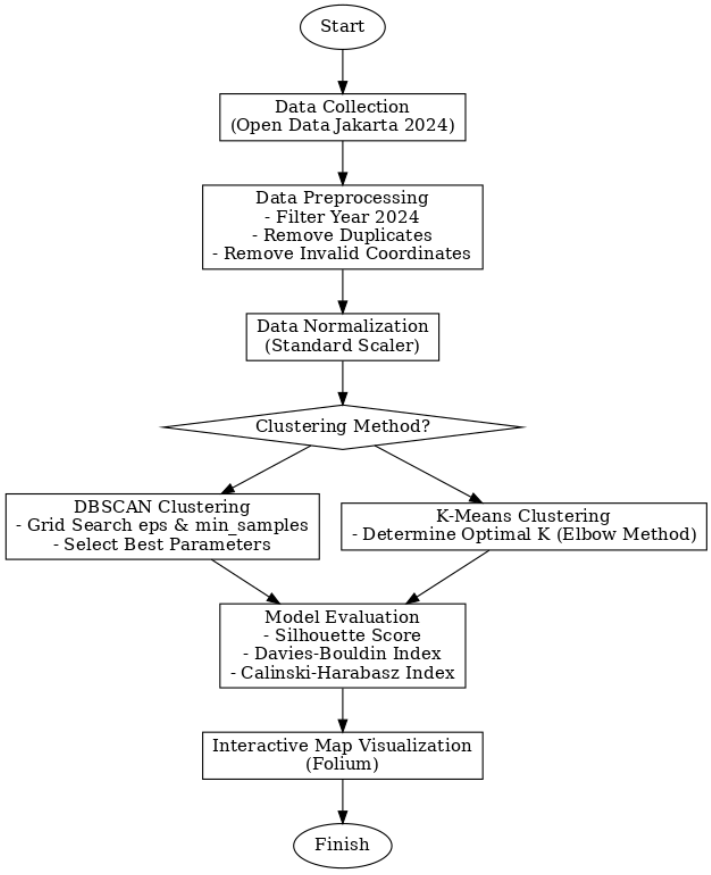


Figure 1. Flowchart of Research Method

Figure 1 shows that the research begins with collecting data on public complaints related to crime in DKI Jakarta in 2024 obtained from the Jakarta Open Data portal. The data used includes information on the location of the incident, type of crime, and geographic coordinates (latitude and longitude). The next stage is data pre-processing which includes filtering data only for the year 2024, removing duplicates, and cleaning data from invalid coordinates. Next, the data was normalized using StandardScaler so that the numerical variables had the same scale before clustering. Clustering analysis was conducted using two methods, namely DBSCAN and K-Means. In DBSCAN, eps and min_samples parameters are optimized through grid search to obtain the best results, while in K-Means the optimal number of clusters is determined using the Elbow Method. The clustering results from both methods are then evaluated using Silhouette Score, Davies-Bouldin Index, and Calinski-Harabasz Index metrics to measure cluster quality and separation. Finally, the clustering results are visualized on an interactive map using Folium, where each cluster is displayed in a different color and the noise points are grayed out, making it easier to identify crime-prone zones in the DKI Jakarta area.

Data Pre-processing

The data pre-processing stage is an important step to ensure data quality and consistency before further analysis is carried out, as emphasized by Nugrohi & Denih (2023) (Nugroho & Denih, 2020). Data that is irrelevant, duplicate, or has invalid values can interfere with the analysis process and reduce the accuracy of the results (Aini & Nasution, 2025).The first step is to filter the dataset to ensure that only crime reports that occurred in the period January-December 2024 are used. This filtering aims to make the analysis reflect current conditions and is relevant to the research objectives. Data selection based on a specific period is a common practice to maintain temporal relevance in spatial data analysis (Sofyan, 2024). Duplicate data were then

removed to avoid bias in the analysis results, especially in density-based methods such as DBSCAN that are sensitive to the number of points in a given area (Simbolon & Friskila, 2024). Duplication can occur due to input errors, redundancy of data from multiple sources, or double recording. This deduplication process improves the accuracy of clustering results and the representation of the distribution of criminal events in the field (Maulana et al., 2025). Furthermore, points that have latitude and longitude coordinate values outside the valid range for the DKI Jakarta area are deleted. Invalid coordinates can be caused by GPS errors, data entry errors, or noise in the location recording process (Nikken et al., 2015). This coordinate validation is part of data cleaning to ensure the accuracy of spatial analysis (Ifrianto, 2020). After the data cleaning process is carried out, the next stage in pre- processing is to normalize the data to ensure format consistency and facilitate analysis in the next stage.

Data Normalization

Data normalization is an important stage before the clustering analysis process, especially in algorithms that are sensitive to scale differences between variables (Farida & Lubis, 2025). In this study, the normalized numerical variables are Latitude and Longitude, which represent the geographical coordinates of the crime scene. The normalization process is performed using the StandardScaler method, which is a standardization technique that changes the distribution of variables to have a mean of 0 and a standard deviation of 1. The following normalization formula (1) is used:

$$Z = \frac{X - \mu}{\sigma} \quad (1)$$

Where X is the original value of the variable, μ is the mean value, and σ is the standard devition. Kalimat perlu di parafrasekan lagi dan lebih diseederhanakan (Kusnaidi et al., 2022). For example, without normalization, differences in longitude degree values can have greater weight than latitude in the Euclidean distance calculation used by algorithms such as K-Means and DBSCAN (Kusnaidi et al., 2022).

Clustering Analysis

Clustering analysis was performed using two methods. First, the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm, which requires *eps* and *min_samples* parameters. These parameters are optimized using *grid search* to get the best *Silhouette Score* value. Second, the K-Means algorithm as a comparison, where the optimal number of clusters is determined using the *Elbow Method*.

DBSCAN Method

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is a density-based clustering algorithm that groups data points based on the number of neighbors within a certain radius(Kong et al., 2019). Points that are in high-density areas will form a cluster, while points in low-density areas are categorized as noise or outliers (Salman, 2023). DBSCAN classifies points into three categories: Core Points, which are points with a minimum number of neighbors min_samples within epsilon distance (eps); Border Points, which are points within eps distance from the core point but have less than min_samples neighbors; and Noise Points, which are points that do not meet the previous two criteria (Alwi & Muliono, 2025). Determining the proximity between points usually uses the Euclidean distance using Formula (2) as follows:

$$d(p,q) = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2} \quad (2)$$

Where p and q are two data points, dan n is the number of dimensions (in this study, 2 dimensions: latitude and longitude).

The two main parameters of DBSCAN are eps, which is the maximum distance between points to be considered neighbors, and min_samples, which is the minimum number of points within the *eps* radius to form a cluster (Willyana et al., 2025), (Wahyuningtyas et al., 2023). Choosing the right parameters is very important: *eps* too small produces a lot of *noise*, while too large can merge different clusters; *min_samples* too small can form clusters from *noise*, while too large makes many points unclustered (Rizki & Sulianta, 2025). As for obtaining optimal parameters, this research applies Grid Search with a predetermined range of *eps* and *min_samples* values.

The DBSCAN model algorithm in this study includes: (1) data pre-processing (filter year 2024, remove duplicates, coordinate validation, normalization), (2) determination of *eps* and *min_samples* ranges, (3) application of DBSCAN for each parameter combination, (4) evaluation using *Silhouette Score*, and (5) selection of the best parameters to form the final cluster which is then visualized on an interactive map. With this optimization, DBSCAN is expected to accurately represent the pattern of crime distribution in DKI Jakarta, as well as identify areas that have a high density of cases and scattered *noise* points.

K-Means Method

K-Means is a partition-based clustering algorithm that divides data into k clusters based on the similarity of the distance between points (Sulistiyawati & Supriyanto, 2020). The main objective is to minimize Within-Cluster Sum of Squares (WCSS), which is the sum of squares of each point's distance to its cluster center (centroid) (Akram et al., 2024). In general, the algorithm works by determining the number of clusters k, selecting an initial point as the centroid (randomly or using k-means++ initialization), clustering points to the cluster with the closest centroid, recalculating the centroid based on the average position of points in each cluster, and repeating the process until the centroid position converges or the maximum number of iterations is reached. The K-Means objective function can be written in Formula (3) as follows:

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n_j} | x_i^{(j)} - \mu_j|^2 \quad (3)$$

where k is the number of clusters, n_j the number of poinst in the-j, $x_i^{(j)}$ the-ii point in the-j, μ_j cluster centroid the-j, and $| x - \mu|$ is the Euclidean distance between the points and the centroid.

The selection of the optimal number of clusters is done using the Elbow Method, which calculates WCSS for various values of *k* and plots the relationship between *k* and WCSS. The *elbow* point on the graph indicates the optimal *k* value, where the decrease in WCSS starts to slow down significantly.

In this study, latitude and longitude variables were used to form spatial clusters of crime in DKI Jakarta. Both variables have been normalized using *StandardScaler* to have an equal scale. Based on the *Elbow Method* results, the optimal k value is obtained which is then used in the K-Means algorithm with *k- means++* initialization. The clustering results produce a *centroid* that represents the average location of criminal incidents in each group, making it easier to identify crime-prone zones. Cluster visualization is done on an interactive map to clearly display the spatial distribution of each group.

Model Evaluation

In this study, the clustering results from the DBSCAN and K-Means methods were evaluated using three evaluation metrics, namely Silhouette Score, Davies-Bouldin Index, and Calinski-Harabasz Index. These three metrics were chosen because they can provide a comprehensive overview of cluster quality, both in terms of density within clusters and separation between clusters

Silhouette Score

Silhouette Score measures how well a point falls within its corresponding cluster compared to other clusters (Hasan, 2024). Its value ranges from -1 to 1, where a value close to 1 indicates that the point fits into its cluster and is well separated from other clusters, a value close to 0 indicates that the point is at the boundary between clusters, while a negative value indicates that the point may be in the wrong cluster; the Silhouette Score formula for a point *i* is written in Formula (4).

$$s(i) = \frac{b(i) - a(i)}{\max\{ a(i), b(i)\}} \quad (4)$$

in this case, $a(i)$ is the average distance of point i to all other points in the same cluster (intra-cluster distance), while $b(i)$ is the average distance of point i to all points in the nearest cluster (nearest-cluster distance).

Davies-Bouldin Index (DBI)

The Davies-Bouldin Index measures the degree of similarity between clusters, where a smaller value indicates better cluster quality (Luthfi & Wijayanto, 2021). The DBI considers the ratio between intra-cluster and inter-cluster distances. The DBI formula is written in Formula (5).

$$DBI = \frac{1}{k}\sum_{i=1}^{k} \max_{j \neq i}\left(\frac{\sigma_i + \sigma_j}{d(c_i, c_j)}\right) \quad (5)$$

In this case, *k* is the number of clusters, σi is the average distance from each point in cluster i to its centroid c_i (intra-cluster distance), and $d(c_i, c_j)$ is the distance between the centroids of clusters i dan j (inter-cluster distance); the lower the DBI value, the better the separation between clusters.

Calinski-Harabasz Index (CHI)

The Calinski-Harabasz Index, also known as the Variance Ratio Criterion, measures the ratio of inter- cluster separation to within-cluster density (Syah et al., 2025). A higher CHI value indicates better cluster quality. The CHI formula is found in Formula (6)

$$CHI = \frac{\mathrm{Tr}(B_k)/(k-1)}{\mathrm{Tr}(W_k)/(n-k)} \quad (6)$$

In this case, B_k is the between-cluster dispersion matrix, W_k is the within-cluster dispersion matrix, *n* is the total number of data points, and k is the number of clusters; this index is effective for assessing the balance between clear cluster separation and high internal density.

RESULTS AND DISCUSSION

Results

The dataset used in this study contains 125 rows and 10 columns containing criminal complaint reports in the DKI Jakarta area in the period January-December 2024. The data is sourced from the Jakarta Open Data portal and contains information related to the identity of

the report, the location of the incident, the time of reporting, and the category of crime. In this study, variable X represents the input feature in the form of geographic coordinates (Latitude and Longitude), while variable Y is the target variable that represents the category of crime (Type_criminal). Initial mapping of incident locations shows that crime hotspots are scattered throughout the administrative area of DKI Jakarta, with a relatively high concentration in Central Jakarta and parts of West Jakarta. The following **Table 1** presents data on crime complaint reports in DKI Jakarta in 2024 used in this study.

Table 1. DKI Jakarta Criminality Complaint Report Data in 2024

Location of Complaint	Latitude	Longitude	...	Type of Crime
Jalan Duri Utara Kelurahan Duri Utara Jakarta Barat	-6.155.454	106.801.807	...	Brawl
Jalan Pusdiklat Depnaker Kelurahan Makasar Jakarta Timur	-6.279.696	106.878.432	...	Brawl
…	…	…	…	…
Jl. Rawa Badung Kelurahan Jatinegara Jakarta Timur	-6,208538294	106,9233617	...	Theft and Mugging

In Table 1, the data includes various locations of criminal incidents throughout DKI Jakarta with a variety of crime categories, such as brawls, theft, sexual harassment, prostitution, and others. This information forms the basis for spatial analysis to identify patterns and crime-prone zones. After the initial understanding of the data, the next step is the data pre-processing stage to ensure quality and consistency before clustering analysis.

Data Pre-processing Result

The data cleaning stage is carried out to ensure that the dataset used is of good quality and ready for analysis. This process includes three main steps, starting with Filtering Data for Year 2024 Only This step aims to make the analysis only focus on criminal events that occur during the period January-December 2024. Filtering is done by utilizing the Period_data column. Then, Deleting Data with Invalid Coordinates.

Data that does not have valid latitude or longitude values will be deleted to avoid errors in spatial analysis. Invalid data can be empty values (NaN), zero, or coordinates outside the geographic range of DKI Jakarta. This step ensures that only points with reasonable and mappable coordinates are retained. After that, duplication removal was done to avoid bias due to double calculations on the same point. Duplications that are removed are rows with all identical column values, so every entry that is the same will be removed, so that each criminal incident in the dataset is only recorded once. After the data is cleaned, data normalization continues to facilitate analysis.

Data Normalization Results

The normalization stage of the criminal type category is carried out to group the various report categories into a more concise and consistent form. This is important to reduce data complexity and facilitate spatial analysis. Based on the initial data, the Type_criminal column contains 29 unique categories with many variations of writing, such as "Theft and Mugging", "Immoral Acts", "Sexual harassment in residential neighborhoods", "Alleged Prostitution", and "Gambling / cockfighting / illegal bird contest". All variations were grouped into seven main categories, namely Brawl, Theft, Sexual Harassment, Alleged Criminal Complaint, Illegal Air Pollution, Prostitution, and Gambling.

This grouping is done using a mapping dictionary in Python to ensure similar categories are in the same group. Thus, the analysis results focus more on regional differences and distribution patterns, not on inconsistent naming variations. Next, data standardization was performed on

e-ISSN: 2722-2691|p-ISSN: 2722-2683

the Latitude and Longitude coordinate variables using the StandardScaler method. This standardization changes the data to a scale with an average of 0 and a standard deviation of 1, so that distance calculations in clustering algorithms such as DBSCAN and K-Means are not biased due to differences in the range of values between variables.

The data normalization and standardization stage was carried out to facilitate analysis. In the normalization of criminal categories, 29 unique categories were merged into 7 main categories, making it easier to group and interpret the data. Furthermore, data standardization is applied to Latitude and Longitude coordinates using StandardScaler, so that the original value with a latitude range of -6.28 to -6.11 and longitude 106.76 to 106.96 is converted into a standardized value with an average of 0 and a standard deviation of 1. Based on the results of normalizing the criminal category, an initial mapping of the point distribution of criminal cases in DKI Jakarta was carried out to see the initial distribution based on the location of the incident. This visualization is presented in Figure 2.

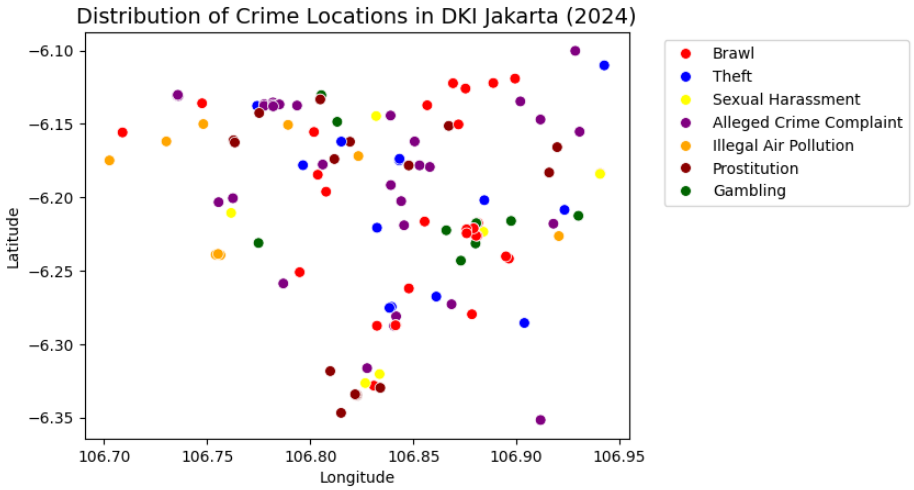


Figure 2. Point Distribution of Criminal Cases in DKI Jakarta

Figure 2 displays a map of the uneven distribution of criminal case points, with a concentration of incidents seen in several areas, such as the city center and densely populated areas. This distribution indicates the potential for criminal hotspots in certain zones.

Prior to the clustering process, all numerical variables were normalized using the StandardScaler method, so that each feature had a mean of zero and a standard deviation of one. This step is important to prevent differences in variable scales from biasing the calculation of spatial distances, which play a crucial role in clustering algorithms such as K-Means and DBSCAN (Kusnaidi et al., 2022). In addition, normalization of the categorical criminal types was performed to reduce data complexity and ensure consistency across variations in category naming. To illustrate this, a comparison graph of the number of categories before and after normalization was created, as shown in Figure 3. This normalization allows the analysis to focus more on spatial distribution patterns rather than inconsistent data representation (Farida & Lubis, 2025).

Figure 3 shows the comparison of coordinate data before and after normalization. After standardization, the distribution of latitude and longitude values has been on a uniform scale, which allows the clustering process to run more optimally.

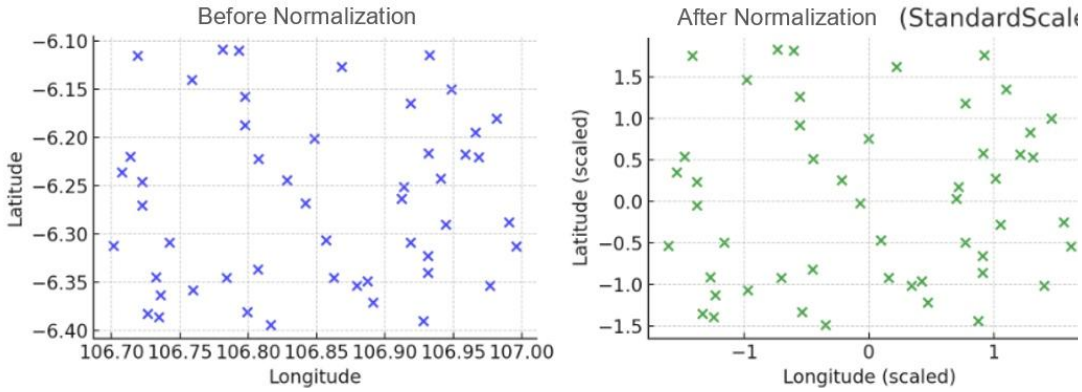


Figure 3. Data Comparison Chart Before and After Data Normalization

DBSCAN Clustering Results

The DBSCAN (Density-Based Spatial Clustering of Applications with Noise) method is applied to cluster crime hotspots in DKI Jakarta based on spatial density. Optimal parameters were obtained through Grid Search with variations of ε = 0.05-0.50 and min_samples = 3-9, evaluated using Silhouette Score. The optimization results showed ε = 0.300 and min_samples = 3, resulting in 12 clusters without noise points. The clustering results are outlined in Figure 4 where the spatial distribution of the clusters is in different colors. Most of the clusters are concentrated in the city center and commercial districts, while other clusters are scattered in areas with fewer crime reports. The absence of noise points indicates that all data have enough spatial proximity to form clusters.

Evaluation of model performance on DBSCAN resulted in a Silhouette Score of 0.302, which indicates the quality of separation between clusters is in the good enough category. The Davies-Bouldin Index value of 1.685 indicates a moderate level of similarity between clusters, while the Calinski-Harabasz Index of 24.285 indicates the ratio of separation between clusters to density within clusters is at a relatively low level, which can be caused by the distribution of points that are quite scattered in some areas.

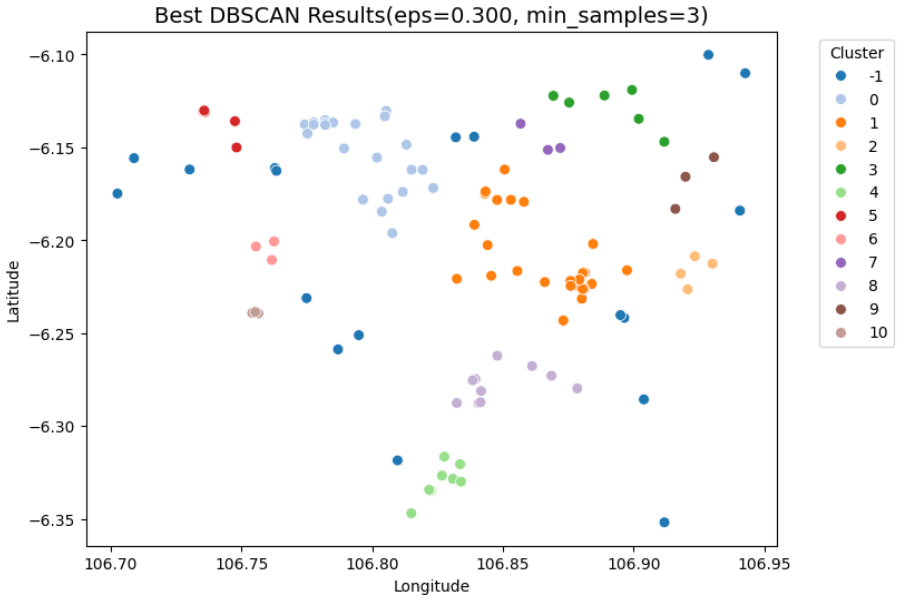


Figure 4. DBSCAN Clustering Results

e-ISSN: 2722-2691|p-ISSN: 2722-2683

Figure 4 shows 12 clusters with shapes that follow the pattern of data density. Clusters tend to form in areas with dense occurrences such as the city center, while peripheral areas have broader or fragmented clusters.

K-Means Clustering Result

As a comparison method to DBSCAN, the K-Means Clustering algorithm is used which works by minimizing the Within-Cluster Sum of Squares (WCSS), so that each point in the cluster has a minimum distance to its cluster center. Determination of the optimal number of clusters is done using the Elbow Method, by plotting the WCSS value against the variation in the number of clusters (k). The Elbow graph can be seen in Figure 5.

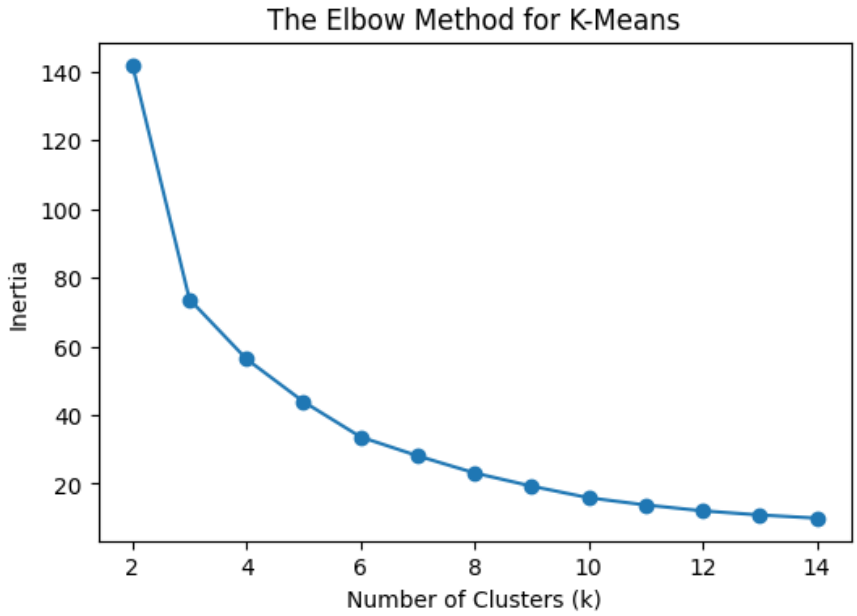


Figure 5. Elbow Method Graph

Figure 5 displays the Elbow graph, which shows the optimal value of k = 3 as the "bending" point where the decrease in WCSS value starts to slow down significantly, indicating the most appropriate number of clusters. In contrast to DBSCAN, K-Means forms spherical clusters with relatively clear boundaries, making it less able to capture complex spatial contours. With k = 3, the K-Means clustering results divide the data into three main clusters whose distribution is quite evenly distributed in the DKI Jakarta area and is visualized in Figure 6.

Figure 6 shows the results of K-Means clustering that divides the DKI Jakarta area into three main clusters. This division is simpler and more organized than DBSCAN, but does not fully follow the natural distribution of crime incidence, as K-Means is more effective on data with isotropic distribution and clear distances between clusters.

Quantitative evaluation showed Silhouette Score = 0.288, indicating a good fit of points to their clusters; Davies-Bouldin Index = 1.781, indicating a slightly higher similarity between clusters than DBSCAN; and Calinski-Harabasz Index = 22.847, indicating the ratio of separation between clusters to internal density was moderate.

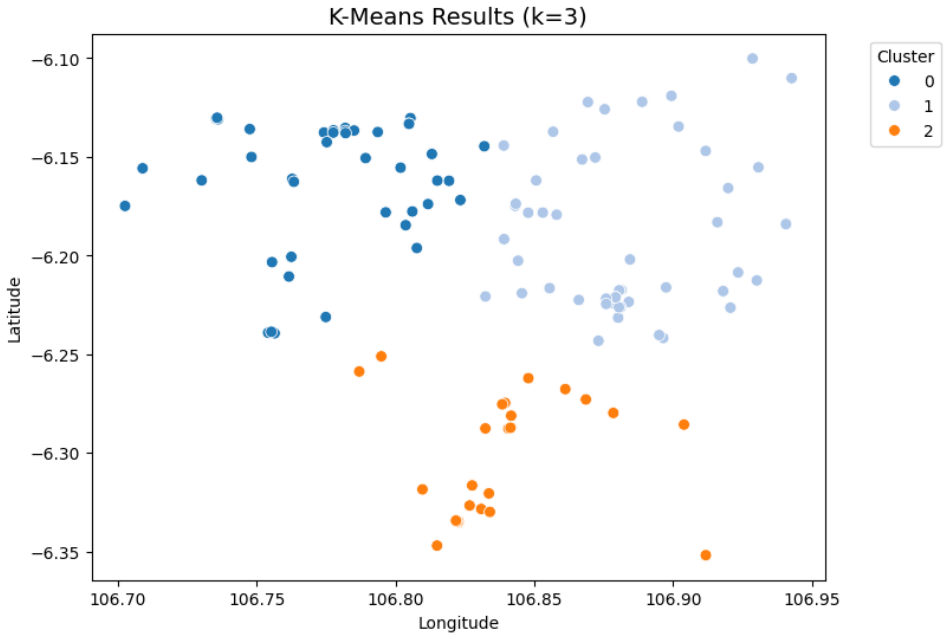


Figure 6. K-Means Clustering Results

Model Evaluation and Comparison

Model Evaluation was conducted by comparing the clustering results of DBSCAN and K-Means using three metrics: Silhouette Coefficient to assess internal cluster uniformity, Davies-Bouldin Index to measure similarity between clusters, and Calinski-Harabasz Index to assess separation between clusters relative to cluster size. The comparison also includes the number of clusters formed as well as the noise points of each method, as presented in Table 2.

Table 2. Comparison of DBSCAN and K-Means Evaluation Results

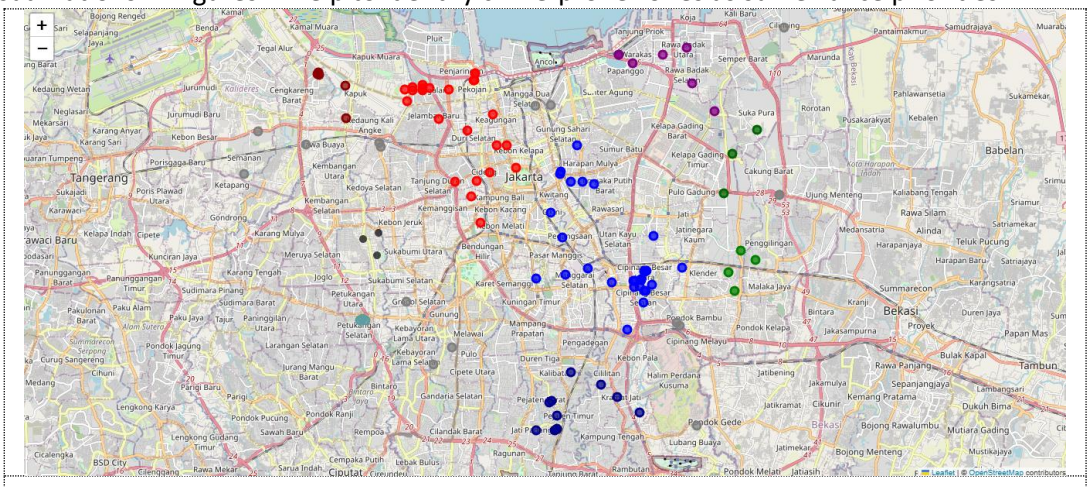
Method	Number of Clusters	Noise	Silhouette	Davies–Bouldin	Calinski–Harabasz
DBSCAN	12	0	0.302	1.685	24.285
K-Means	3	0	0.483	0.725	145.897

Table 2 shows that the K-Means analysis results in more compact and clearly separated clusters, indicatedby Silhouette = 0.483, Davies-Bouldin = 0.725, and Calinski-Harabasz = 145.897, so that the global separation of regions is more optimal. Meanwhile, DBSCAN formed more clusters (12), captured spatial variations in detail, and detected hotspots and small-scale risk zones according to point density, although the quantitative metrics were lower. Spatially, DBSCAN excels in providing a realistic depiction of risk maps, while K-Means is more suitable for structured segmentation of administrative areas. Thus, the two methods complement each other. DBSCAN is more effective for detecting density-based patterns and micro-hotspots because it can identify irregular cluster shapes and handle noise or outliers, while K-Means is more suitable for structured segmentation at the macro level, particularly when clusters are relatively homogeneous and well-separated.

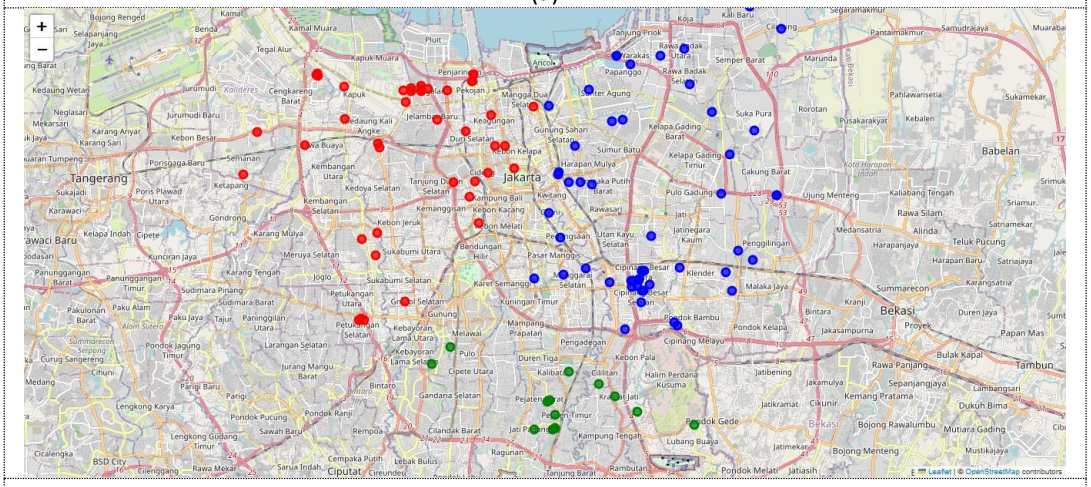
Visualization of Result on Interactive Maps

The results of DBSCAN and K-Means clustering were visualized using Folium to display the spatial distribution of criminal cases in DKI Jakarta. In DBSCAN, each cluster is given a different color, while noise points are marked gray, so that areas with high density can be clearly seen. K-Means

also marks the clusters with different colors, but all points fall into one of the clusters without any noise, resulting in a more even distribution but less sensitive to variations in density. The visualizations in Figures 7 help to identify crime-prone zones as surveillance priorities.



(a)



(b)

Figure 7. Interactive map of DBSCAN clustering results (a) and Interactive map of K-Means clustering results (b).

Figure 7 shows the results of K-Means clustering on crime locations in DKI Jakarta 2024, with each color representing an evenly distributed cluster, forming large zones in Central, East, and South Jakarta. There are no noise points, but the clusters are round and follow the natural density pattern less than DBSCAN

Discussion

The results show that K-Means forms clusters that are more compact, clearly separated, and have large inter-cluster variation, indicated by higher Silhouette values, lower Davies-Bouldin Index, and higher Calinski-Harabasz Index than DBSCAN. In contrast, DBSCAN formed more clusters and was able to detect noise points, making it more sensitive to variations in case density and able to identify small-scale crime hotspots. This finding is in line with previous research, such as the study by (Kong et al., 2019) and recent applications of DBSCAN in Indonesia (2022), which also found that DBSCAN excels in detecting areas of high incidence density and local patterns not seen in K-Means. The implication is that DBSCAN is more appropriate for

detailed mapping of crime-prone zones, while K-Means is more suitable for macro mapping that supports the division of security forces' working areas.

Furthermore, the crime patterns identified in this study indicate a concentration of incidents at certain points, which are strongly related to population density, economic activity, and accessibility (Zhang et al., 2020; Li & Cheng, 2023). These findings provide a practical foundation for designing targeted interventions, such as deploying more security personnel in high-density clusters, installing surveillance systems in identified hotspots, or improving street lighting and public facilities in vulnerable areas. On a macro level, K-Means results can be used to allocate resources and define police jurisdiction boundaries more efficiently, ensuring that high-crime zones receive proportional attention. Therefore, clustering-based spatial crime analysis not only reveals patterns but also supports evidence-based policy and preventive strategies for urban crime management (Putra et al., 2021; Andini et al., 2022).

CONCLUSION

This research analyzed the spatial distribution of criminal cases in DKI Jakarta 2024 using DBSCAN and K-Means. K-Means formed more compact and separated clusters, with Silhouette (0.483), Calinski- Harabasz Index (145.897), and Davies-Bouldin Index (0.725) showing good cluster quality. DBSCAN formed more clusters (12) and was able to detect noise points, excelling in capturing density variations and micro-hotspots. The contribution of this study lies in the comparison of two clustering methods for recent location-based crime data at the provincial scale, emphasizing that DBSCAN is suitable for detailed risk mapping, while K-Means is more suitable for macro mapping of security forces' working areas. The findings can serve as a reference for the development of spatial analysis models and data-driven security policies.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to the course lecturer for the guidance, valuable suggestions, and constructive feedback provided throughout the completion of this research paper. The authors also thank Universitas Ahmad Dahlan for providing academic support during this study.

DECLARATION

Author contribution

All authors contribute in the research and/or writing the paper, and approved the final manuscript.

Sintia Afriyani	Conceptualized the research, designed the methodology, collected and curated the data, performed the survival analysis, interpreted the results, and prepared the original manuscript.
Winarsi J. Bidul	Assisted in developing the research methodology, validated the data analysis, contributed to the interpretation of results, and reviewed the manuscript.
Imam Al Maksur	Assisted in data analysis, reviewed the research findings, contributed to manuscript revision, and improved the English writing of the manuscript.
Achmad Fauzan	Supervised the research, provided methodological guidance, critically reviewed the manuscript, and contributed to the interpretation of the results.
Roza Azizah Primatika	Provided academic supervision, reviewed and revised the manuscript, validated the research methodology, and approved the final version of the manuscript.

Funding

This research did not receive any funding.

Conflict of interest

All authors declare that they have no competing interests.

Ethics declaration
We as authors acknowledge that this work has been written based on ethical research that conforms with the regulations of our institutions and that we have obtained the permission from the relevant institutes when collecting data. We support the Bulletin of Applied Mathematics and Mathematics Education (BAMME) in maintaining the high standards of personal conduct, practicing honesty in all our professional practices and endeavors.

The use of artificial intelligence
The authors used AI tools only for brainstorming research ideas. No AI-generated content was included in the manuscript, and all analyses and conclusions were performed by the authors.

Additional information
Not available.

REFERENCES

Aini, N., & Nasution, M. I. P. (2025). *Akurasi kualitas data informasi pada sistem manajemen nurul*. *2*(1), 40–50.

Akram, A., Risal, N., Surianto, D. F., Teknik, F., Makassar, U. N., Ekonomi, I. K., Maju, D., & Berkembang, D. (2024). *Mini-batch K-Means clustering untuk pengelompokan*. 235–244.

Alwi, B., & Muliono, R. (2025). *Analisis klustering menggunakan algoritma DBSCAN untuk deteksi anomali dalam data transaksi keuangan*.

Farida, J. I., & Lubis, A. H. (2025). *Pengelompokan lokasi pariwisata di indonesia dengan menggunakan variasi distance pada algoritma K-Means*.

Hasan, Y. (2024). *Pengukuran Silhouette Score dan Davies-Bouldin Index pada Hasil Cluster*. *06*(01), 60–74.

Ifrianto, T. (2020). *Validasi Lokasi Perizinan Perkebunan Dalam Pengendalian Pemanfaatan Ruang di Kabupaten Paser*. *3*(2), 109–117.

Julian, Y. A., & Umar, G. (2025). *Peran Sistem Tata Kelola Lingkungan Berbasis data Spasial dalam Perencanaan*. *2*(1), 1–8.

Kong, Q., Yang, J., Yuan, L., & Yang, J. (2019). Use Density-Based Spatial Clustering of Applications with Noise (DBSCAN) Algorithm to Identify Galaxy Cluster Members. *Earth and Environmental Science PAPER*. https://doi.org/10.1088/1755-1315/252/4/042033

Kusnaidi, M. R., Gulo, T., & Aripin, S. (2022). *Penerapan Normalisasi Data Dalam Mengelompokkan Data Mahasiswa Dengan Menggunakan Metode K-Means Untuk Menentukan Prioritas Bantuan Uang Kuliah Tunggal*. *3*(4), 330–338. https://doi.org/10.47065/josyc.v3i4.2112

Luthfi, E., & Wijayanto, A. W. (2021). *dalam pengelompokkan indeks pembangunan manusia Indonesia Comparative analysis of hirearchical , k-means , and k-medoids clustering and methods in grouping Indonesia ' s human development index*. *17*(4), 761–773.

Maulana, D. I., Andang, A., Usrah, I., & Purnomo, A. (2025). *Deteksi Duplikasi Data pada Sistem Pemantauan Kualitas Udara Berbasis IoT*. *14*, 138–144. https://doi.org/10.22146/jnteti.v14i2.16272

Miftahurrahmi, S., Amalita, N., & Mukhti, T. O. (2024). *DBSCAN Method in Clustering Provinces in Indonesia Based on Crime Cases in 2022*. *2*(2010), 330–337.

Muin, A., Rakuasa, H., Geografi, M. P., Jakarta, U. N., Geografi, D., & Indonesia, U. (2023). *Pemetaan Kerentanan Kebakaran Hutan di Pulau Buru , Provinsi Maluku Berdasarkan Fire Hotspot*. *2*(4), 675–683. https://doi.org/10.55123/insologi.v2i4.2256

Munajat, A. A., & Yusuf, H. (2024). *Faktor yang mempengaruhi tingkat kejahatan di kota besar dynamics of urban criminality : a study of the factors affecting crime rates in large cities*. 1330–1339.

Nikken, D., Sirin, S., Salyasari, N. D., & Maryanto, A. (2015). *Standardisasi Prosedur Pengambilan Foto Udara dengan Pesawat LSA untuk Pengembangan Payload Inderaja*. *1*.

Nugroho, B., & Denih, A. (2020). Perbandingan kinerja metode pra-pemrosesan. *Jurnal Ilmiah IImu Komputer Dan Matematika*, *17*(2), 381–387.

Rizki, M. F., & Sulianta, F. (2025). *Analisis Klasterisasi Data Peserta Asuransi PT Xyz Menggunakan Metode Density-Based Spatial Clustering of Applications with Noise (DBSCAN)*. 993–1001. https://doi.org/10.33364/algoritma/v.22-1.2417

Salman, N. (2023). *DENSITY-BASED CLUSTERING ANALYSIS*. *8*, 1–8.

Septianto, M. A., Faqih, A., & Rinaldi, A. R. (2025). *Klasterisasi data produksi pertanian di kabupaten cirebon dengan algoritma K-*. *13*(2).

Simbolon, I. N., & Friskila, P. D. (2024). *Analisis dan evaluasi algoritma dbscan spatial clustering of applications with noise) pada tuberkulosis*. *12*(3).

Sofyan, L. P. (2024). *Analisis determinan stunting di kabupaten bogor dan kota bogor : pendekatan spasial untuk meningkatkan efektivitas intervensi*.

Stiawan, B., & Yusuf, H. (2025). *Dampak Kriminalitas Terhadap Kualitas Hidup Masyarakat Urban*. *2*(6), 308–312.

Sulistiyawati, A., & Supriyanto, E. (2020). *Implementasi Algoritma K-means Clustring dalam Penetuan Siswa Kelas Unggulan*. *15*(2), 25–36.

Syah, F., Wiyono, P., Kaffi, L., & Maulana, M. H. (2025). *Segmentasi Wilayah Provinsi di Indonesia Berdasarkan Indeks Penanganan Stunting Menggunakan PCA dan Partition Clustering*. *2025*(Senada), 406–417.

Syahri, R., Informatika, T., & Selatan, S. (2023). *Algoritma K-Means clustering : sebuah studi literatur K-Means clustering algorithm : a literatur study*. *x*(x), 1–7. https://doi.org/10.12345/juri

Wahyuningtyas, F. D., Arafat, A., Stiawan, A., & Rolliawati, D. (2023). *DBSCAN pada Analisis Data Penjualan Melalui Facebook*. *14*(1), 7–16.

Willyana, I. P., Hadiana, A. I., & Ilyas, R. (2025). *Jurnal Analisis Klaster Daerah Rawan Gempa di Indonesia Cluster Analysis of Earthquake Prone Areas in Indonesia*. *12*(1), 59–71.

Yusuf, H., Zanudin, S., & Karno, U. B. (2025). *Pengaruh faktor sosial ekonomi terhadap tingkat kriminalitas di kota metropolitan*. *2*(2), 2505–2516.

e-ISSN: 2722-2691|p-ISSN: 2722-2683

This page is intentionally left blank.