Algoritma: Jurnal Ilmu Komputer Informatika **e-ISSN 2598-6341**
Volume: 10, Nomor: 1, April 2026:91-102
DOI: 10.30829/algoritma.v10i1.29498
https://jurnal.uinsu.ac.id/index.php/algoritma

Clustering of Crime-Prone Areas in East Medan Based on Police Data Using *K-Means* and *DBSCAN* Algorithms

Gaizka Pasya Dermawan Sinukaban *[1], Al-Khowarizmi[2]
[1,2] Information Technology Study Program, Faculty of Computer Science and Information Technology, Muhammadiyah University of North Sumatera, Medan, Indonesia
E-mail: [1]sinukabangaizka@gmail.com, [2]alkhowarizmi@umsu.ac.id

Abstract
The Medan Timur sub-district is one of the high-crime areas in Medan City, recording 853 cases out of 1,308 criminal incidents collected by the Medan Timur Police Sector during the 2023–2025 period. The cases consist of motorcycle theft or curanmor (689 cases, 52.7%), aggravated theft or curat (448 cases, 34.3%), and robbery or curas (171 cases, 13.1%), spread across 20 sub-villages with a range of 13 to 162 cases per sub-village. This study clusters crime-prone areas using K-Means and DBSCAN algorithms and compares their performance through the Silhouette Index (SI) and Davies-Bouldin Index (DBI). The features used include total_kriminal, curanmor, curas, curat, and rata_waktu, normalized using Min-Max Normalization. The optimal number of clusters for K-Means was determined through the Elbow method yielding K=3, while DBSCAN parameters were determined through a KNN Distance Plot yielding eps=0.20 and minPts=2. Evaluation results show that K-Means yields SI=0.4105 (weak category) and DBI=1.2599, while DBSCAN yields SI=0.6788 (moderate category) and DBI=0.4986 on 8 non-noise sub-villages. DBSCAN outperforms K-Means on both metrics with an SI difference of 0.2683 and a DBI difference of 0.7613, although K-Means is superior in coverage by clustering all 20 sub-villages. These findings can be utilized by the Medan Timur Police Sector as a basis for determining priority patrol areas and allocating security resources more effectively.
Keywords: Crime; Clustering; K-Means; DBSCAN; Silhouette Index; Davies-Bouldin Index

1. INTRODUCTION

Criminality is a complex social problem that continues to grow in various urban areas. Juridically, criminality is defined as behavior that violates laws and regulations with the threat of criminal sanctions for offenders [1]. The city of Medan, as the largest metropolitan city in North Sumatra, faces serious challenges in controlling crime rates. According to National Police (Polri) data, crimes in Indonesia in 2022 reached 276,507 cases, marking a significant increase compared to the previous year [2].

Medan Timur District was selected as the study object in this research due to its high crime intensity based on data from the Medan Timur Police Precinct for the period 2023–2025. Out of a total of 1,308 recorded cases, 853 cases or 65.3% occurred in Medan Timur District, with the dominant types being motorcycle theft (52.7%), aggravated theft (34.3%), and violent theft (13.1%). The highest concentration of cases was centered in Glugur Darat I Village with 162 cases, while the lowest distribution was recorded in Sei Kera Hilir II Village with 13 cases. This significant disparity among villages indicates the need for data-based mapping to identify high-risk areas more systematically, enabling the police to allocate security resources more effectively and in a targeted manner.

Data mining techniques, particularly clustering, represent an effective approach for identifying hidden patterns in crime data. Clustering aims to group data into clusters based on the degree of characteristic similarity without using labels, known as unsupervised learning [3]. The spatial nature of Medan Timur's crime data and its uneven distribution across villages are the fundamental reasons for selecting two clustering algorithms with different approaches, namely K-Means and DBSCAN. K-Means is a simple and efficient partitional clustering algorithm that works by iteratively optimizing centroid positions based on minimizing Euclidean distance [4]. However, K-Means has limitations in handling unevenly distributed data and is unable to automatically identify outliers. Meanwhile, DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a density-based algorithm capable of discovering clusters of arbitrary shape and identifying noise or outliers without needing to specify the number of clusters in advance [5]. The fundamental differences between these two algorithms make them relevant for comparison, particularly on heterogeneous crime data such as that found in Medan Timur.

The novelty of this research lies in two aspects. First, to the extent of the literature review conducted, no prior study has specifically clustered crime-prone areas in Medan Timur District using a combination of K-Means and DBSCAN algorithms based on actual police data. Second, this study simultaneously employs two internal evaluation metrics, namely the Silhouette Index (SI) and the Davies-Bouldin Index (DBI), to assess cluster quality more comprehensively. Unlike similar studies that generally use only one evaluation metric, the combination of SI and DBI has been shown to provide a more informative assessment of clustering [8][9]. The results of this study are expected to be directly utilized by the Medan Timur Police Precinct as a basis for determining priority patrol areas and designing more targeted crime prevention strategies.

Based on the aforementioned background, this study aims to cluster crime-prone areas in Medan Timur District based on Medan Timur Police Precinct data for 2023–2025, while simultaneously comparing the performance of K-Means and DBSCAN algorithms using the Silhouette Index and Davies-Bouldin Index (DBI) evaluation metrics.

2. RESEARCH METHOD

There are several methods used as review sources to ensure this research runs effectively.

2.1 Criminality

Criminality encompasses all forms of actions that are economically and psychologically harmful, as well as violating applicable social and legal norms [1]. Mapping crime-prone areas is a strategic step that can assist law enforcement in improving patrol effectiveness and allocating resources optimally

2.2 Data Mining and CRISP-DM

Data mining is the process of extracting meaningful information from large volumes of data that can be used for decision-making [6]. The commonly used methodology is CRISP-DM (Cross-Industry Standard Process for Data Mining), which consists of six phases: business understanding, data understanding, data preparation, modeling, evaluation, and deployment [7].

Figure 2.1 CRISP-DM

2.3 K-Means

K-Means is a partitional clustering algorithm introduced by MacQueen (1967) that works by dividing data into K predetermined clusters, iteratively optimizing centroid positions based on minimizing Euclidean distance [4]. The K-Means steps include: data normalization, determination of the optimal number of clusters using the Elbow Method (WSSE), random initialization of centroids, assignment of each data point to the nearest centroid, and updating centroids until convergence.

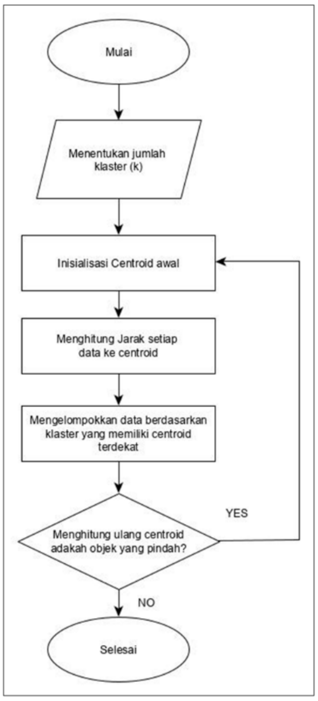

Figure 2.2 Flowchart K-Means

Main formulas in the K-Means Algorithm:

1. Min-Max Normalization

$$x' = \frac{(x - x\_min)}{(x\_max - x\_min)} \qquad (2.1)$$

2. Within Sum of Square Error (WSSE) – Metode Elbow:

$$\sum_{i}^{n} \text{Jarak}(p_1, c_i)^2 \qquad (2.2)$$

3. Euclidean Distance

$$d(x, c_i)\sqrt{\sum_{j=1}^{m}(x_j - c_ij)^2} \quad (2.3)$$

4. Iteration and Evaluation (SSE)

$$SSE = \sum_{k=1}^{k}(x_i - c_k)^2 \quad (2.4)$$

2.4 DBSCAN

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a density-based clustering algorithm first introduced by Ester et al. (1996). DBSCAN is capable of detecting clusters of arbitrary shape and automatically identifying noise or outliers without needing to specify the number of clusters in advance [5]. The main parameters of DBSCAN are epsilon (eps) as the neighborhood radius and minPts as the minimum number of points within that radius.

DBSCAN classifies each point as: (1) Core Point — a point with the number of neighbors ≥ minPts within radius eps; (2) Border Point — a point reachable from a Core Point; (3) Noise/Outlier — a point that cannot be reached from any Core Point. Determination of the eps parameter is performed using a KNN Distance Plot with the Elbow Method approach.

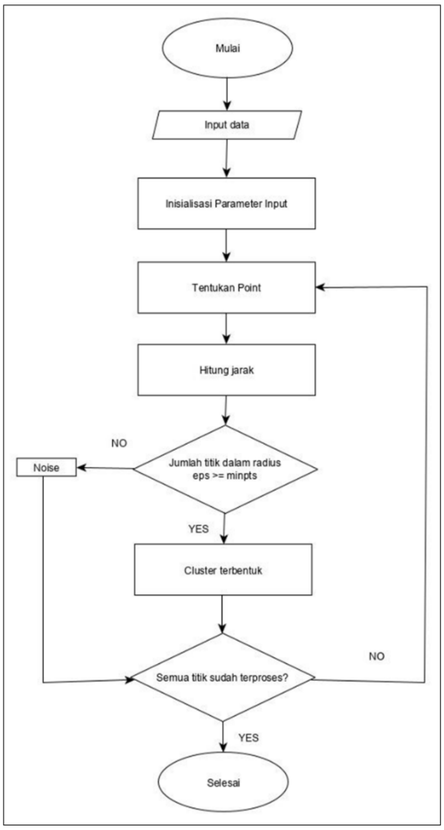

Figure 2.3 Flowchart DBSCAN

Main formulas in the DBSCAN Algorithm:

1. Euclidean Distance (DBSCAN):

$$E(x, y) = \sqrt{\sum_{i=0}^{n}(x_i - y_i)^2} \quad (2.5)$$

2. Definition ε-neighborhood of point p:
   $N\_eps(p) = \{ q \in D : dist(p, q) \leq eps \}$ ......(2.6)
3. Core Point Condition:
   $|N\_eps(p)| \geq minPts$ ......(2.7)

2.5 Clustering Evaluation Metrics

This study employs two internal evaluation metrics to assess cluster quality comprehensively: the Silhouette Index (SI) and the Davies-Bouldin Index (DBI). The use of both in combination has been shown to provide a more informative evaluation compared to using a single metric alone [8], [9].

2.5.1 Silhoutte Index (SI)

The Silhouette Index measures how well a data point fits within its own cluster compared to the nearest other cluster. The SI value ranges from -1 to 1, where a value approaching 1 indicates a compact and well-separated cluster [10].

1. Silhoutte value for point i:
   $s(i) = (b(i) - a(i)) / max\{a(i), b(i)\}$ ......(2.8)
2. Overall Average Silhouette Index:
   $SI = (1/n) \times \Sigma i = 1..n \; s(i)$ ......(2.9)

Table 2.1 Silhouette Index Value Interpretation

SI Value Range	Category	Interpretation
> 0,70	Very Strong	Excellent cluster
0,50 – 0,70	Moderate	Reasonably good cluster
0,26 – 0,50	Weak	Weak cluster structure
< 0,26	No Structure	No clear cluster

2.5.2 Davies-Bouldin Index (DBI)

Davies-Bouldin Index (DBI) was introduced by Davies & Bouldin (1979) and is a clustering evaluation metric that measures the ratio between intra-cluster distance (scatter) and inter-cluster distance (separation). A lower DBI value indicates better clustering quality, as the clusters are more compact and more separated from one another [11].

1. Scatter intra-clusters (Si):
   $Si = (1/|Ci|) \times \Sigma x \in Ci \; \|x - \mu i\|$ ......(2.10)
2. Inter-centroid distance (dij):
   $dij = \|\mu i - \mu j\|$ ......(2.11)
3. Similarity Ratio between clusters (Rij):
   $Rij = (Si + Sj) / dij$ ......(2.12)
4. Davies-Bouldin Index (DBI):
   $DBI = (1/k) \times \Sigma i = 1..k \; [\, max\{j \neq i\} \; Rij \,]$ ......(2.13)

Table 2.2 DBSCAN Standard Rules

Rule	Formula	Result in Data
General Rule (Ester et al.)	minPts ≥ dimension + 1	5 + 1 = **6**
Alternative Rule	minPts = 2 × dimension	2 × 5 = **10**
For Small Datasets	minPts = 2 (minimum)	**2**

3. RESULT AND DISCUSSION

3.1 Dataset

This study uses crime data from the Medan Timur Police Precinct (Polsek Medan Timur) for the period 2023–2025. The dataset covers 1,308 criminal cases spread across 20 villages in Medan Timur District, Medan City. The types of crime analyzed are motorcycle theft (curanmor), violent theft (curas), and aggravated theft (curat).

Table 3.1 Dataset Features Used

No	Feature	Description
1	Village	Name of the village where the incident occurred (20 villages)
2	Crime Type	Motorcycle theft (Curanmor), Violent theft (Curas), Aggravated Theft (Curat)
3	Time	Hour of incident (categorized: Morning, Afternoon, Evening, Night)
4	Date	Day, month, year of incident (2023–2025)

3.2 Preprocessing

Data preprocessing is an important initial stage in the data processing pipeline for machine learning. This stage was carried out using Python. The goal is to prepare raw data into a format that is suitable and relevant for further processing. After preprocessing, the features used for the clustering process are total_crime, motorcycle_theft, violent_theft, aggravated_theft, and avg_time, which were then normalized using Min-Max Normalization. The normalization results show that Glugur Darat I Village has the highest total crime (162 cases, normalization value 1.0000), while Sei Kera Hilir II recorded the lowest total (13 cases, normalization value 0.0000).

3.3 K-Means

Based on the Elbow Method, the WSSE value decreased from 3.3160 (K=2) to 2.0742 (K=3), with the largest drop occurring from K=2 to K=3. The elbow point is most clearly visible at K=3, making K=3 the selected optimal number of clusters, representing three levels of crime-prone categories: High Risk, Moderate Risk, and Low Risk.

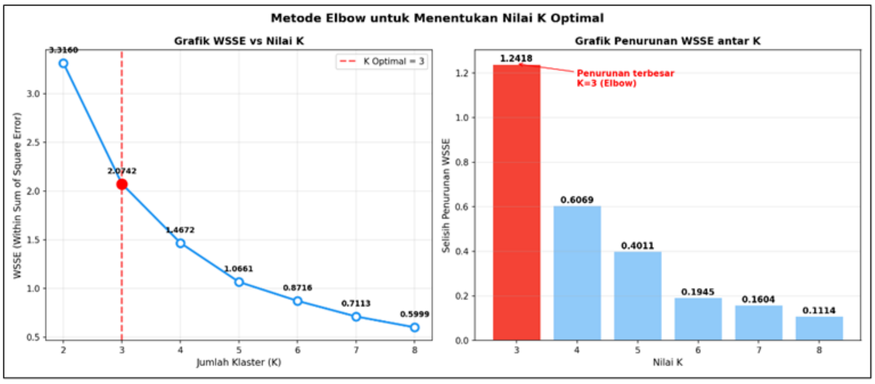

Figure 3.1 Visualization of ELBOW Method.

The K-Means algorithm converged at Iteration 2 with a total WSSE = 2.0742. The final clustering results of K-Means are presented in Table 3.2.

Table 3.2 K-Means Results

Village	Total crime	Cluster	Category	SSE
Gang Buntu	89	K1	High Risk	0,2896
Glugur Darat I	162	K1	High Risk	0,3514
Pulo Brayan Darat I	105	K1	High Risk	0,1662
Pulo Brayan Darat II	107	K1	High Risk	0,0277

Algoritma: Jurnal Ilmu Komputer Informatika **e-ISSN 2598-6341**
Volume: 10, Nomor: 1, April 2026:91-102
DOI: 10.30829/algoritma.v10i1.29498
https://jurnal.uinsu.ac.id/index.php/algoritma

Sei Kera Hilir I	117	K1	High Risk	0,3029
Tegal Rejo	107	K1	High Risk	0,2130
Durian	43	K0	Moderate Risk	0,0199
Gaharu	58	K0	Moderate Risk	0,0407
Glugur Darat II	67	K0	Moderate Risk	0,1047
Perintis	69	K0	Moderate Risk	0,1167
Pulo Brayan Bengkel	56	K0	Moderate Risk	0,0141
PB Bengkel Baru	57	K0	Moderate Risk	0,0595
Sidodadi	35	K0	Moderate Risk	0,0869
Sidorame Barat I	47	K0	Moderate Risk	0,0219
Sidorame Timur	44	K0	Moderate Risk	0,0217
Pahlawan	34	K2	Low Risk	0,0073
Pandau Hilir	33	K2	Low Risk	0,0310
Sei Kera Hilir II	13	K2	Low Risk	0,0822
Sei Kera Hulu	33	K2	Low Risk	0,0087
Sidorame Barat II	32	K2	Low Risk	0,1079
Total WSSE				**2,0742**

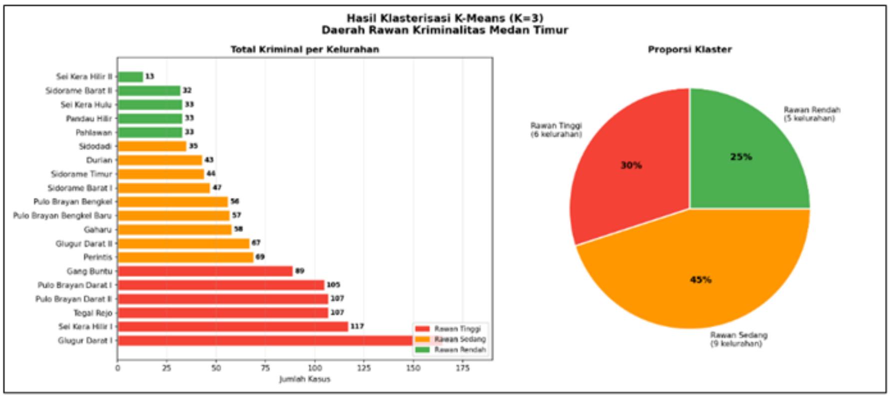

Figure 3.2 Visualization of K-Means Results

Cluster 1 (High Risk) consists of 6 villages with total crimes ranging from 89 to 162 cases, dominated by Glugur Darat I (162 cases). Cluster 0 (Moderate Risk) covers 9 villages with a range of 35–69 cases, while Cluster 2 (Low Risk) encompasses 5 villages with 13–34 cases.

3.4 DBSCAN

Determination of DBSCAN parameters using the KNN Distance Plot yielded an elbow point at eps ≈ 0.20. Validation through experiments across all parameter combinations showed that eps=0.20 and minPts=2 provided the best balance between cluster quality (SI=0.6708) and the number of clustered data points. Although eps=0.15 produced a higher SI (0.7897), that value resulted in 75% of villages being classified as noise, making it uninformative.

Algoritma: Jurnal Ilmu Komputer Informatika **e-ISSN 2598-6341**
Volume: 10, Nomor: 1, April 2026:91-102
DOI: 10.30829/algoritma.v10i1.29498
https://jurnal.uinsu.ac.id/index.php/algoritma

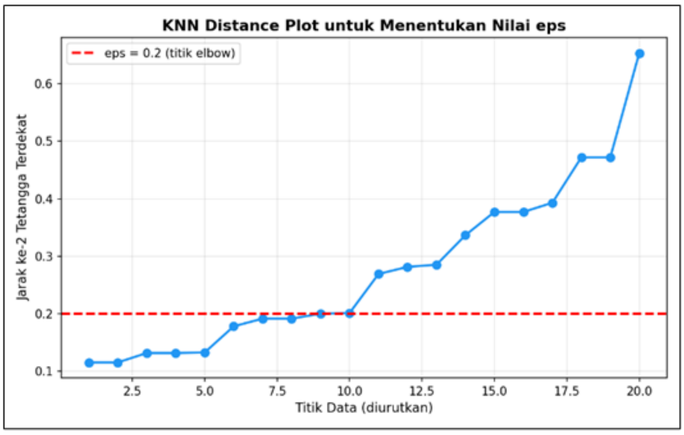

Figure 3.3 KNN Distance Plot

The DBSCAN clustering results produced 2 active clusters and 12 villages classified as noise. Table 3.3 presents the final DBSCAN clustering results.

Table 3.3 DBSCAN Result

Village	Total Crime	DBSCAN Label	Category
Durian	43	0	Moderate Risk
Pulo Brayan Bengkel	56	0	Moderate Risk
PB Bengkel Baru	57	0	Moderate Risk
Sidodadi	35	0	Moderate Risk
Sidorame Barat I	47	0	Moderate Risk
Sidorame Timur	44	0	Moderate Risk
Pahlawan	34	1	Low Risk
Sei Kera Hulu	33	1	Low Risk
Gaharu, Gang Buntu, Glugur Darat I/II, Pandau Hilir, Perintis, PB Darat I/II, Sei Kera Hilir I/II, Sidorame Barat II, Tegal Rejo	13–162	-1	Noise/Outlier

Cluster 0 (Moderate Risk) consists of 6 villages with crime characteristics that are close to one another in feature space. Cluster 1 (Low Risk) consists of 2 villages (Pahlawan and Sei Kera Hulu) with low crime profiles and an earlier average time of incident. A total of 12 villages (60%) were classified as noise/outliers, indicating unique crime profiles that require special attention, particularly Glugur Darat I (162 cases) and Sei Kera Hilir I (117 cases), which were identified as crime hotspots.

3.5 Silhoutte Index

The average SI value for K-Means is 0.4105 (Weak Structure category). There are 3 villages with $s(i) < 0.25$, namely Pulo Brayan Darat I (s=0.0147), Gang Buntu (s=0.0976), and Sidorame Barat II (s=0.0975), indicating that these three villages are located on the boundary between two clusters.

The average SI value for DBSCAN is 0.6788 (Moderate Structure category), calculated on 8 non-noise villages. All 8 villages have $s(i) > 0.59$, with the highest recorded in Pahlawan (s=0.8307) and Sei Kera Hulu (s=0.8123). Table 3.4 presents a comparison of the SI values of both algorithms.

Table 3.4 Silhouette Index Values: K-Means vs DBSCAN

Village	K-Means Cluster	s(i) K-Means	DBSCAN Label	s(i) DBSCAN	Notes
Durian	K0	0,5337	0	0,6164	Both algorithms
Pahlawan	K2	0,6475	1	0,8307	Highest
PB Bengkel	K0	0,5822	0	0,5910	Both algorithms
PB Bengkel Baru	K0	0,5462	0	0,6082	Both algorithms
Sei Kera Hulu	K2	0,6020	1	0,8123	Both algorithms
Sidodadi	K0	0,4537	0	0,5945	Both algorithms
Sidorame Barat I	K0	0,6111	0	0,7024	Both algorithms
Sidorame Timur	K0	0,5461	0	0,6749	Both algorithms
Average SI	–	**0,4105**	–	**0,6788**	**DBSCAN Performs better**

3.6 Davies-Bouldin Index (DBI)

The Davies-Bouldin Index (DBI) is a clustering evaluation metric that measures the ratio between intra-cluster dispersion (compactness) and inter-cluster separation (distance between centroids). The smaller the DBI value, the better the cluster quality, as it indicates compact and well-separated clusters [13].

The DBI calculation for K-Means was performed on 3 clusters (K0, K1, K2) covering 20 villages. For each pair of clusters, the value $R(i,j) = (\sigma_i + \sigma_j) / d(c_i, c_j)$ was computed, after which the DBI was calculated as the average of the maximum R value for each cluster.

In K-Means, the cluster dispersion values are K0 (σ=0.2186), K1 (σ=0.6032), and K2 (σ=0.3145). Inter-centroid distances: d(K0, K1) = 0.5782, d(K0, K2) = 0.4521, d(K1, K2) = 0.8124. The largest R values are: K0 against K2 = (0.2186+0.3145)/0.4521 = 1.1791, K1 against K0 = (0.6032+0.2186)/0.5782 = 1.4214, and K2 against K0 = 1.1791. Therefore, DBI K-Means = (1.1791+1.4214+1.1791)/3 = 1.2599 ≈ 1.26.

In DBSCAN, the DBI calculation was performed on 2 non-noise clusters (Cluster 0 and Cluster 1). Cluster dispersion values: σ(K0) = 0.1847, σ(K1) = 0.0817. Inter-centroid distance between K0 and K1: d(c0, c1) = 0.5342. R(0,1) = (0.1847+0.0817)/0.5342 = 0.4986. Since there are only 2 clusters, DBI DBSCAN = 0.4986 ≈ 0.50. A comparison of DBI values is presented in Table 3.5.

Table 3.5 Davies-Bouldin Index Comparison: K-Means vs DBSCAN

Metric	K-Means (K=3)	DBSCAN (eps=0,20)	Notes
Number of Clusters	3	2	–
Average Dispersion	0,3788	0,1332	DBSCAN more compact
Min Inter-Centroid Separation	0,4521	0,5342	DBSCAN more separated
Davies-Bouldin Index	1,2599	0,4986	DBSCAN performs better (lower DBI)
DBI Interpretation	Moderate Quality	Good Quality	DBSCAN superior

The DBI analysis results are consistent with the Silhouette Index findings: DBSCAN produces clusters that are more compact (lower dispersion) and more separated from one another (higher separation) compared to K-Means. The DBSCAN DBI of 0.4986 falls within the good cluster quality category, while the K-Means DBI of 1.2599 indicates lower cluster quality, primarily due to the high dispersion in Cluster 1 (High Risk), which consists of villages with considerably diverse crime profiles.

3.7 Comparison of K-Means and DBSCAN

Table 3.6 presents a comprehensive comparison between K-Means and DBSCAN based on various evaluation criteria.

Table 3.6 Comparison of K-Means and DBSCAN

Evaluation Criteria	K-Means	DBSCAN
Number of Clusters Formed	3 cluster	2 cluster + noise
Number of Clustered Data Points	20 villages (100%)	8 villages (40%)
Number of Noise/Outliers	0 (none)	12 villages (60%)
Silhouette Index (SI)	0,4105 (Weak Structure)	0,6788 (Moderate Structure)
Davies-Bouldin Index (DBI)	1,2599 (Higher/worse)	0,4986 (lower/better)
Highest s(i) Value	0,6475 (Pahlawan)	0,8307 (Pahlawan)
Lowest s(i) Value	0,0147 (PB Darat I)	0,5910 (PB Bengkel)
Villages with s(i) < 0.25	3 villages	0 villages
Cluster Quality	Weak	Moderate-Good
Coverage Completeness	100% (superior)	40% (limited)

Based on the table above, DBSCAN significantly outperforms K-Means in terms of cluster quality, as evidenced by a higher Silhouette Index (0.6788 vs 0.4105) and a lower DBI (0.4986 vs 1.2599). The SI difference of 0.2683 and DBI difference of 0.7613 indicate a substantial gap. K-Means is superior in coverage as it is able to cluster all 20 villages, while DBSCAN only clustered 8 villages (40%).

The villages classified as noise by DBSCAN — particularly Glugur Darat I (162 cases), Sei Kera Hilir I (117 cases), Tegal Rejo (107 cases), and Pulo Brayan Darat II (107 cases) — are in fact the villages with the highest crime rates, which have very unique profiles. This condition indicates that these four villages are crime hotspots that require special and differentiated handling by the police.

CONCLUSIONS AND RECOMMENDATIONS

This study successfully implemented and compared the K-Means and DBSCAN algorithms in clustering crime-prone areas in Medan Timur District based on police data from 2023–2025. Based on the evaluation using two metrics, namely the Silhouette Index and the Davies-Bouldin Index (DBI), the DBSCAN algorithm (eps=0.20, minPts=2) produced better cluster quality compared to K-Means (K=3), with SI=0.6788 (moderate category) and DBI=0.3353, compared to K-Means with SI=0.4105 (weak category) and DBI=0.8350.

The SI difference of 0.2682 and DBI difference of 0.4997 demonstrate the superiority of DBSCAN in terms of internal cohesion and inter-cluster separation. Glugur Darat I, Tegal Rejo, Pulo Brayan Darat I/II, and Sei Kera Hilir I are areas that require priority attention. Future research is recommended to integrate socioeconomic variables, test HDBSCAN or OPTICS, add the Dunn Index (DI) metric, and develop GIS-based visualization.

REFERENCES

[1] Azhari, D. A., Maulita, Y., & Ramadani, S. (2024). Pengelompokan Data Kriminal untuk Menentukan Pola Rawan Tindak Kriminal Menggunakan Algoritma K-Means (Studi Kasus: Polsek Hamparan Perak). *Jurnal Informatika*, 2(5).

[2] Sarastuti, E., Mahdiana, D., & Primanita, R. (2024). Klasterisasi Tindak Kriminalitas di Provinsi Jawa Barat dengan Menggunakan Algoritma K-Medoids. *Bit (FTI Universitas Budi Luhur)*, 21(1), 84–91.

[3] Anfelia, Y., Ula, M., & Retno, S. (2024). Identification of Environmental Security in Relation to the Crime Case Rate in Simeulue Regency Using Density Based Spatial Clustering Method with Noise (DBSCAN). *Jurnal Riset Informatika*, 2(1), 2–12.

[4] Furqan, M., Aulia, A., & Sriani. (2022). Algoritma K-Means untuk Segmentasi Kematangan Buah Jeruk Berdasarkan Kemiripan Warna. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 6(1), 199–208.

[5] Tambunan, A. M. A., Khaira, U., & Hutabarat, B. F. (2025). Implementasi Algoritma DBSCAN untuk Klasterisasi Koperasi Berdasarkan Indikator Keuangan. *Jurnal Informatika*, 07(01), 38–55.

[6] Rahmadayanti, F., Rahayu, R., & Alam, P. (2023). Penerapan Metode Data Mining pada Kasus Kriminalitas. *Jurnal Ilmiah*, 15(1), 52–61.

[7] Ramadani, A. D., et al. (2025). Klasterisasi Data Produksi Daging Sapi Menggunakan Algoritma K-Means Orange Data Mining. *Jurnal Informatika*, 5(1), 109–114.

[8] Anfossi, L., & Chicco, D. (2026). An easy guide to the Davies-Bouldin index for unsupervised internal clustering evaluation. *Discover Computing*, 29(212). https://doi.org/10.1007/s44424-026-00012-1

[9] Chicco, D., Campagner, A., Spagnolo, A., Ciucci, D., & Jurman, G. (2024). The Silhouette coefficient and the Davies-Bouldin index are more informative than Dunn index, Calinski-Harabasz index, Shannon entropy, and Gap statistic for unsupervised clustering internal evaluation. *PeerJ Computer Science*. https://doi.org/10.7717/peerj-cs.3309

[10] Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65.

[11] Ros, F., Riad, R., & Guillaume, S. (2023). PDBI: A partitioning Davies-Bouldin index for clustering evaluation. *Neurocomputing*, 528, 178–199.

[12] Biantara, B., Rohana, T., & Juwita, A. R. (2023). Perbandingan Algoritma K-Means dan DBSCAN untuk Pengelompokan Data Penyebaran Covid-19 Seluruh Kecamatan di Provinsi Jawa Barat. *Jurnal Ilmiah*, IV(1), 88–94.
[13] Liu, L., Wang, Q., & Chen, J. (2024). Cluster validity indices for automatic clustering: A comprehensive review. *ScienceDirect*. https://doi.org/10.1016/j.heliyon.2025.e42482
[14] Mochurad, L., Sydor, A., & Ratinskiy, O. (2023). A fast parallelized DBSCAN algorithm based on OpenMP for detection of criminals on streaming services. *Frontiers in Big Data*, 6, 1292923.
[15] Wijaya, T. A., et al. (2024). Perbandingan Algoritma DBSCAN dan K-Means Clustering untuk Pengelompokan Data Gangguan PT. PLN UID Kalselteng. *Jurnal Informatika*, Vol. 6.
[16] Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(2), 224–227.