BITNET
Jurnal Pendidikan Teknologi Informasi
Vol. 11 No. 3, September 2026, pp.34-40
https://journal.umpr.ac.id/index.php/bitnet/issue/...
DOI: https://doi.org/...

Implementation of K-Means for Mapping Cleanliness Levels in Pancur Batu District

Raisya Dwi Zahra Putri[1], Mansur AS[2], Hamidah Nasution[3], Insan Taufik[4], Adidtya Perdana[5]
[1,2,3,4,5]Medan State University, Indonesia
zahraraisya86@gmail.com

ABSTRACT

This study aims to classify areas in Pancur Batu District based on environmental cleanliness levels using the K-Means Clustering method, analyze the characteristics of each cluster based on cleanliness parameters, and develop a web-based mapping system to visualize the clustering results spatially. The study utilized 578 image data analyzed using five parameters, namely visible waste density, drainage condition, percentage of clean roads, percentage of households with trash bins, and waste collection frequency. The results showed that the areas in Pancur Batu District were successfully grouped into three clusters: Cluster 0 (Clean) consisting of 6 villages, Cluster 1 (Dirty) consisting of 12 villages, and Cluster 2 (Very Clean) consisting of 7 villages. The clustering results were then visualized through a web-based mapping system to facilitate users in identifying the distribution of cleanliness levels across the study area. Expert validation conducted by a geography specialist obtained a total score of 42 with an average score of 4.2, which falls into the good category. Therefore, the implementation of the K-Means Clustering method was able to effectively classify areas based on environmental cleanliness levels and provide useful information to support environmental cleanliness management and decision-making processes in Pancur Batu District.

Keywords: K-Means Clustering, environmental cleanliness, regional mapping, web-based system.

Article history

Received:	*Revised:*	*Accepted*	*Published:*
15 July 2026	*10 August 2026*	*21 August 2026:*	*3 September 2026*

INTRODUCTION

In recent years, the issue of environmental hygiene has become a crucial concern in various regions across Indonesia, driven by increasing community activities and the resulting volume of waste (Subekti & Prayoga, 2021). Environmental hygiene is fundamentally understood as a condition that reflects a region's level of waste management, sanitation, and the prevention of pollution, particularly from solid waste. Effective waste management is essential, as poor practices lead to significant waste accumulation, which degrades environmental quality and poses a threat to public health (Rahmadhani et al., 2023). A clean and well-maintained environment is vital because it has a direct positive impact on the health, comfort, and overall quality of life for the community (Pohan et al., 2023).

However, Pancur Batu District faces significant challenges in environmental management, characterized by visible waste piles in public spaces and malfunctioning drainage systems. Ideally, environmental management should be supported by accurate and structured data; yet, the reality shows that local authorities face difficulties in planning and prioritizing interventions due to the lack of measured data regarding hygiene levels across its villages (Tarigan et al., 2023; Ritonga & Fedryansyah, 2022). This situation necessitates a systematic, data-driven approach to transform complex location-based data into visual information that stakeholders can easily understand. Mapping plays a pivotal role in this process, serving as a medium to visualize phenomena within a specific area (Rozak, 2021). The process involves collecting, analyzing, and interpreting geospatial data to produce informative representations of real-world conditions (Gunawan & Sutrisno, 2021).

Systematic representation of the earth's surface through maps allows users to understand patterns, distributions, and relationships within geospatial data (Komalasari, 2023). Specifically, the use of thematic maps is highly effective for presenting specific variables to facilitate spatial analysis and location-based decision-making (Gaffara et al., 2024). Through such visualizations, complex data can be simplified and made more intuitive for users (Mita Wahyuni, 2025). In this study, environmental hygiene is measured through five primary parameters: visible waste density, drainage condition, waste collection frequency, percentage of clean roads, and the percentage of households with trash bins.

These parameters are highly relevant as they directly represent physical aspects and waste management efficiency in residential areas. Improved household sanitation practices and waste management at the source, such as the provision of closed trash bins, contribute significantly to the overall improvement of the residential environment (Muzaqi dkk., 2025). Furthermore, the regularity of waste collection is essential to prevent accumulation that can pollute soil and waterways (Muzaqi dkk., 2025). To process these quantitative parameters, the K-Means Clustering method is employed to group areas based on their similar characteristics (Yuan & Yang, 2023). This algorithm works by minimizing the distance between data points and their respective centroids to achieve high homogeneity within groups (Jain, 2021).

Before the clustering process begins, data normalization is a critical stage to standardize the scales of different parameters, ensuring that each variable contributes equally to the calculation (Syamsiah et al., 2025). The results of this classification are then implemented into a web-based system using the **Python** programming language (Mei Hellyana et al., 2023). Spatial visualization in this system is not only presented numerically but also through more intuitive geographic representations using polygons to show the administrative boundaries and coverage of each village (Novianti et al., 2021). The primary objective of this research is to classify areas in Pancur Batu District into dirty, clean, and very clean categories to support targeted decision-making in environmental hygiene management.

METHOD

This research employs a descriptive quantitative approach utilizing clustering analysis to examine the distribution of environmental hygiene levels across Pancur Batu District. In line with the principles of systematic mapping, this study involves the collection, analysis, and interpretation of geospatial data to transform complex information into an easily understood visual medium (Gunawan & Sutrisno, 2021). The primary subjects of analysis are the 25 villages within the district, which are treated as geographical units with specific hygiene characteristics.

The research evaluates these villages based on five quantitative parameters: visible waste density, waste collection frequency, drainage condition, percentage of clean roads, and the percentage of households with closed trash bins. Data collection was conducted through direct field observations, interviews, and documentation. Field observations involved assessing visible waste density on a scale of 0–3 and drainage functionality on a scale of 0–2. Interviews with local officials and community members were specifically used to determine waste collection frequency, which was converted into a scale from 0 to 3. To ensure representative household-level data, a sample of 20 houses per village was observed to calculate the percentage of households utilizing closed trash bins.

For the data analysis, the raw quantitative values first underwent Z-score normalization using the *StandardScaler* function from the *scikit-learn* library. This step is critical to standardize parameter scales and prevent bias in distance calculations, ensuring each variable contributes equally to the clustering process (Syamsiah et al., 2025). The K-Means Clustering algorithm was then implemented to group the villages based on similarity, aiming to minimize the Euclidean distance between data points and their respective centroids (Yuan & Yang, 2023). The optimal number of clusters was determined through the Elbow method and Silhouette coefficient, which identified k=3 as the most stable value for categorizing villages into dirty, clean, and very clean groups.

The final stage involved the implementation of an interactive Web-based Geographic Information System (WebGIS) developed using the Python programming language with the Streamlit framework and Folium library (Mei Hellyana et al., 2023). Following the advice of spatial experts, the system utilizes administrative shapefiles (.shp) to represent each village as a polygon rather than a point, providing a more accurate representation of administrative boundaries and coverage (Novianti et al., 2021). To ensure scientific validity, the methodology and mapping outputs were subjected to expert validation by a geography and GIS specialist using a five-point Likert scale instrument covering ten distinct assessment indicators.

FINDINGS AND DISCUSSION

This sections presents the findings obtained from the implementation of the K-Means Clustering algorithm for classifying environmental cleanliness levels in Pancur Batu District. The clustering process was performed using five environmental cleanliness parameters after data normalization with the StandardScaler method. The optimal number of clusters was identified through the Elbow Method and Silhouette Coefficient. The clustering results were subsequently visualized using a WebGIS application to facilitate spatial interpretation. The following subsections describe the clustering results,parameter characteristics of each cluster, and the discussion of the findings in comparison with previous studies.

Findings

The environmental cleanliness data collected from 25 villages in Pancur Batu District were analyzed using the K-Means Clustering algorithm. Before clustering, all variables were normalized using the StandardScaler method to eliminate differences in measurement scales among the five parameters, namely visible waste density, waste collection frequency, drainage condition, percentage of clean roads, and percentage of households with waste bins. The optimal number of clusters was determined using the Elbow Method and Silhouette Coefficient, resulting in three clusters.

Table 1. Distribution of Villages by Cluster

Cluster	Category	Number of Villages
0	Clean	6
1	Dirty	12
2	Very Clean	7

Table 1 presents the distribution of villages based on the K-Means clustering results. A total of 25 villages were successfully classified into three environmental cleanliness categories. Cluster 1 (Dirty) consists of 12 villages, making it the largest cluster, followed by Cluster 2 (Very Clean) with 7 villages and Cluster 0 (Clean) with 6 villages. This distribution indicates that environmental cleanliness conditions are not uniform across Pancur Batu District and demonstrates the ability of the K-Means algorithm to distinguish villages based on similarities in the selected environmental cleanliness parameters.

The characteristics of each cluster were further analyzed by calculating the average value of each environmental parameter.

Table 2. Average Value of Each Parameter for Each Cluster

Parameter	Clean (Cluster 0)	Dirty (Cluster 1)	Very Clean (Cluster 2)
Visible Waste Density	1,69	2,09	1,44
Waste Collection Frequency	1,5	2,08	1,57

Drainage Condition	0,98	0,94	0,57
Percentage of Clean Roads	1,08	0,38	0,38
Percentage of Houses with Waste Bins	0,29	0,34	0,55

Table 2 presents the average value of each environmental cleanliness parameter for the three clusters. Each cluster exhibits distinct characteristics based on the five evaluated parameters. Cluster 1 (Dirty) has the highest average visible waste density (2.09), indicating poorer environmental cleanliness. Although this cluster also records the highest waste collection frequency (2.08), the percentage of clean roads remains low (0.38), suggesting that frequent waste collection alone is insufficient to maintain overall environmental cleanliness.

Cluster 0 (Clean) represents villages with moderate environmental conditions. This cluster has a relatively lower visible waste density (1.69) than the Dirty cluster and records the highest percentage of clean roads (1.08). However, the drainage condition is relatively less favorable (0.98), indicating that improvements are still needed in drainage management.

Cluster 2 (Very Clean) demonstrates the best overall environmental characteristics. It has the lowest visible waste density (1.44), the best drainage condition (0.57), and the highest percentage of households with waste bins (0.55). These findings indicate that villages in this cluster maintain cleaner environmental conditions and better waste management practices compared to the other clusters.

To improve data interpretation, the clustering results were integrated into a WebGIS application that visualizes the environmental cleanliness categories spatially.

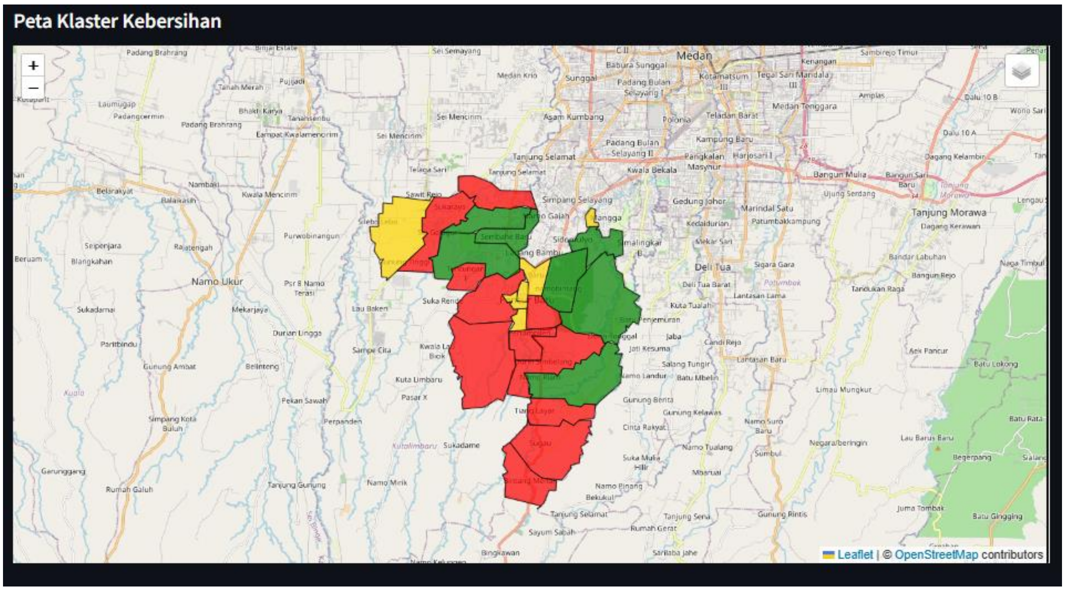

Figure 1. WebGIS Visualization of Enviromental Cleanliness Clusters in Pancur Batu District

Figure 1 presents the spatial distribution of environmental cleanliness across Pancur Batu District. Each village is displayed as a polygon and assigned a different color according to its cluster category. The WebGIS also provides interactive pop-up information containing the environmental parameter values for each village, enabling users to identify cleanliness conditions more efficiently.

Discussion

The findings demonstrate that the K-Means clustering algorithm effectively classified villages with similar environmental cleanliness characteristics into three distinct clusters. The classification was based on five environmental parameters, allowing villages with comparable characteristics to be grouped objectively according to their similarity. This result indicates that environmental cleanliness in Pancur Batu District varies considerably among villages and cannot be accurately represented by a single indicator. Therefore, the combination of multiple environmental parameters provides a more comprehensive assessment of environmental cleanliness.

The characteristics of each cluster reveal different levels of environmental cleanliness. Villages in the Dirty cluster generally exhibit higher visible waste density and lower environmental quality, indicating that waste management practices still require significant improvement. Meanwhile, villages in the Very Clean cluster demonstrate lower waste density, better drainage conditions, and a higher proportion of households with waste bins, reflecting more effective environmental management. The Clean cluster represents intermediate conditions where several environmental indicators are relatively good but still require further improvement to achieve the characteristics of the Very Clean cluster.

The findings of this study are consistent with previous studies that demonstrated the effectiveness of the K-Means algorithm for regional clustering based on environmental and waste management indicators. Magriaty, Murtilaksono, and Anwar (2023) applied the K-Means algorithm to identify waste management zones in Tapin Regency, Indonesia. Their study successfully classified villages into three management zones, enabling the government to prioritize waste management services based on regional characteristics. Similarly, Mad'hika and Pirman (2025) employed K-Means clustering to segment waste management performance across Indonesian provinces and concluded that the clustering results could support evidence-based environmental policy and strategic planning. These findings are consistent with the present study, where K-Means successfully classified villages in Pancur Batu District into three environmental cleanliness categories based on five environmental cleanliness parameters. The clustering results provide objective information that can assist local governments in identifying priority areas for improving environmental cleanliness.

Compared with previous studies, the present research provides an additional contribution by integrating K-Means clustering with an interactive WebGIS platform. While Magriaty et al. (2023) focused on identifying waste management zones for planning purposes and Mad'hika and Pirman (2025) concentrated on clustering provincial waste management performance, this study extends these approaches by visualizing the clustering results spatially through WebGIS. The developed system enables users to identify the environmental cleanliness category of each village directly on an interactive map while simultaneously displaying detailed information on the five environmental cleanliness parameters. This integration not only improves the interpretation of clustering results but also provides a practical decision-support tool that can assist local governments in environmental monitoring, planning, and policy implementation.

CONCLUSION

The implementation of the K-Means Clustering method successfully addressed the critical need for an objective, data-driven assessment of environmental hygiene in Pancur Batu District, as initially proposed in the research objectives. While the study began with the challenge of overcoming subjective and unstructured cleanliness data, the results have effectively bridged this gap by categorizing the 25 villages into three distinct, measurable clusters: Cluster 0 (Clean) with 6 villages, Cluster 1 (Dirty) with 12 villages, and Cluster 2 (Very Clean) with 7 villages. This outcome demonstrates high compatibility between the idealized goal of systematic regional mapping and the actual findings, providing a clear spatial representation of the district's sanitation landscape.

Beyond merely identifying clusters, this research offers a substantialization of meaning by transforming quantitative parameters into a functional decision-support tool. The integration of clustering results into a web-based Geographic Information System (WebGIS) allows local authorities to move beyond general observations and instead implement targeted, priority-based interventions. Specifically, the identification of Cluster 1 as the most critical area provides a transparent roadmap for the local government to allocate resources toward improving waste collection frequency and drainage maintenance where they are most needed.

To further enrich these findings, future research should explore the potential of incorporating more dynamic parameters, such as community participation levels and daily waste generation volumes. Developing the system into a real-time monitoring platform would also enhance its utility, ensuring that environmental management in Pancur Batu remains responsive to the evolving conditions of its residential and public spaces.

REFERENCES

Gaffara, G. R., Aryaguna, P. A., Kurniawan, S., & Nurhaidar, W. O. (2024). PEMETAAN TEMATIK DESA BERBASIS SISTEM INFORMASI GEOGRAFIS. Tata Kota Dan Daerah, 16(2), 179–184. https://doi.org/10.21776/ub.takoda.2024.016.02.6

Gunawan, W., & Sutrisno, B. (2021). Pemetaan sosial untuk perencanaan pembangunan masyarakat. scholar.archive.org. In Sawala: Jurnal pengabdian. https://scholar.archive.org/work/tx6go4ls6rcnfeovp7jpk2evwa/access/way back/http://jurnal.unpad.ac.id/sawala/article/download/32761/pdf

Jain, A. K. (2021). Data clustering: 50 years beyond K-means. Pattern Recognition Letters, 31(8), 651–666.

Komalasari, M. D. (2023). Pemetaan kebutuhan belajar peserta didik dalam pembelajaran berdiferensiasi. Prosiding Seminar Nasional PGSD UST. https://jurnal.ustjogja.ac.id/index.php/sn-pgsd/article/view/15116

Mei Hellyana, C. (2023). Implementasi Bahasa Phyton Untuk Clusterisasi Data Penjualan Menggunakan Metode K-Means. In Informatics and Computer Engineering Journal (Vol. 3, Issue 2). https://jurnal.bsi.ac.id/index.php/ijec/

Muzaqi, L., Istiklaili, F., Khakim, R. L., Arkan, M. Z. R., Ayuningsih, Z., Pertiwi, K. D., & Wirdati, I. E. (2025). Penguatan Sanitasi dan Pengelolaan Limbah Rumah Tangga di Kelurahan Kalisegoro Semarang. INDONESIAN JOURNAL OF COMMUNITY EMPOWERMENT (IJCE), 7(2), 214-220.

Novianti, S., Nurkholifa, T., Suryana, M., Susanto Program Studi Usaha Perjalanan Wisata, E., & Administrasi Niaga Politeknik Negeri Bandung, J. (2021). Penggunaan Geographical Information System (GIS) untuk Visualisasi Analisis Perilaku Spasial Wisatawan. 4(2), 2654–3894 https://doi.org/10.17509/jithor.v4i2,%20October.37168.

Pohan, M., & Rialdy, N. (2023). Upaya Peningkatan Kebersihan Lingkungan, Perilaku Hidup Bersih dan Sehat dalam Peningkatan Kesejahteraan Rakyat. Jurnal Pengembangan Dan Pengabdian Masyarakat Multikultural, 1(2), 71 77. https://doi.org/10.57152/batik.v1i2.782

Rahmadhani, D., Nuraini, C., Abdiyanto, A., Sugiarto, A., & Millanie, F. (2023). RANCANGAN PENGELOLAAN KEBERSIHAN LINGKUNGAN DI KOTA PEMATANG SIANTAR. ARMADA : Jurnal Penelitian Multidisiplin, 1(12), 1408–1414 https://doi.org/10.55681/armada.v1i12.1079.

Ritonga, M. A. P., & Fedryansyah, M. (2022). Konflik Agraria: Perampasan Tanah Rakyat Oleh Ptpn Ii Atas Lahan Adat Masyarakat (Studi Kasus Desa Launch, Simalingkar a, Kecamatan Pancur Batu, Jurnal Kolaborasi Penelitian Ilmiah. academia.edu. https://www.academia.edu/download/104585239/39993-161796-3-PB.pdf

Rozak, I. (2021). Analisis Dan Perancangan Sistem Informasi Geografis Pemetaan Hama Tanaman Padi. Jurnal Informatika Dan Rekayasa Perangkat Lunak. https://jim.teknokrat.ac.id/index.php/informatika/article/view/1239

Subekti, S., Prayoga, I., & Sudrajat, A. S. E. (2020). Pemberdayaan Masyarakat Dalam Pengelolaan Sampah Sebagai Upaya Penanganan Kawasan Kumuh Di Kawasan Pecangaan Kabupaten Jepara. Indonesian Journal of Spatial Planning, 1(2), 45-50.

Syamsiah, N. O., Purwandani, I., Rosmiati, M., & Nurwahyuni, S. (2025). Optimalisasi Algoritma K-Means Melalui Normalisasi Rasio dan Pendekatan SHapley Additive exPlanations (SHAP). Jurnal SITECH : Sistem Informasi Dan Teknologi, 8(2), 233–240 https://doi.org/10.24176/sitech.v8i2.16253.

Tarigan, A. R., Lubis, Z., & Syarifah, S. (2023). Pengaruh pengetahuan, sikap dan dukungan keluarga terhadap diet hipertensi di desa Hulu Kecamatan Pancur Batu tahun 2016. Jurnal Kesehatan. https://journal3.uin-alauddin.ac.id/index.php/kesehatan/article/view/5107

Wahyuni, M. (2025). PEMANFAATAN DATA SPASIAL DALAM PEMETAAN TEMATIK DI PADUKUHAN PROMASAN, KABUPATEN KULONPROGO. JURNAL MULIA, 4(1), 197-204. https://doi.org/10.47002/jpm.v4i1.912

Yuan, C., & Yang, H. (2023). Research on K-value selection method of K-means clustering algorithm. In Journal Automation Computer. mdpi.com. https://www.mdpi.com/2571-8800/2/2/16