Journal of Information Systems Engineering and Management
2025, 10(24s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Optimization of Environmentally Based Waste Management Strategy in Indonesia Using Machine Learning

Andi Riansyah [1,2*], Purwanto Purwanto[1] and Rahmat Gernowo[1]
[1] *Doctoral Program of Information System, Universitas Diponegoro, Semarang, Indonesia*
[2] *Department of Informatics, Universitas Islam Sultan Agung, Semarang, Indonesia*

ARTICLE INFO

Received: 21 Dec 2024

Revised: 09 Feb 2025

Accepted: 26 Feb 2025

ABSTRACT

Introduction: Waste management is one of the biggest environmental challenges facing Indonesia today. With a population of over 270 million people spread across 34 provinces, the country produces a significant amount of waste every day. Diversity in population density, economic activity, and urban development across provinces results in variations in waste production patterns and composition. Effective waste management is critical not only for environmental sustainability, but also for public health and economic development. However, the lack of waste management strategies tailored to the unique characteristics of each province often results in inefficient waste handling and disposal.

Objectives: To classify Indonesian provinces into different clusters based on waste production, volume, and composition using machine learning algorithms. Analyze the characteristics of each cluster to understand the unique challenges and opportunities in waste management they face. Provide recommendations for waste management policies tailored to the specific needs of each cluster.

Methods: The study began with data collection from official sources such as the Ministry of Environment and Forestry or the Central Statistics Agency. After the data was collected, data preprocessing was carried out to clean the data from missing values and outliers, and to normalize the data so that all variables have the same scale. Next, a clustering algorithm such as K-Means was chosen to group provinces based on their waste characteristics. The optimal number of clusters was determined using the Elbow Method

Results: The clustering results divided the provinces into several groups. Cluster 1 contains provinces with relatively low to medium daily waste volumes. Cluster 2 includes provinces with medium waste volumes, while Cluster 3 consists of provinces with very high waste production. The majority of provinces are in clusters 1 and 2, indicating that only a few regions have major problems in waste management.

Conclusions: This study shows that clustering with K-Means can help understand waste production patterns in various provinces in Indonesia. Provinces with similar waste characteristics are grouped into three main clusters, with the majority in the low to medium waste volume category. It was found that organic waste is more dominant than inorganic waste, especially in areas with high waste production. This shows that waste management strategies based on recycling, composting, and renewable energy can be effective solutions. The results of this clustering can be used as a basis for designing more appropriate waste management policies, both in increasing waste processing capacity and encouraging community participation in reducing waste.

Keywords: Waste Management, Environmental Sustainability, Machine Learning, K-Means Clustering, Waste Classification

INTRODUCTION

Optimizing environmental-based waste management strategies is a major challenge in Indonesia, given its rapid population growth and urbanization. With more than 270 million people spread across 34 provinces, the country

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

produces a large volume of waste every day, which if not managed properly can lead to environmental pollution and ecosystem degradation [1], [2]. The lack of strategies based on regional characteristics and suboptimal management infrastructure often hamper the effectiveness of waste management systems [3], [4]. Therefore, a technology-based approach is needed to improve the efficiency of waste management systems and support sustainability policies [5], [6].

Efficient waste management requires a method that is able to group areas or management units based on the capacity and type of waste produced [7], [8]. The application of Machine Learning, especially K-Means Clustering, can be used to classify waste production and distribution patterns based on various variables, such as waste volume, type of waste, and management effectiveness [9], [10]. This approach allows for more accurate mapping of the potential, challenges, and effectiveness of waste management systems, thus supporting more precise data-based decision making [11].

The application of K-Means Clustering in optimizing waste management strategies allows for more structured segmentation based on production patterns and waste processing levels in various regions [12]. Previous studies have shown that this method can improve the optimization of resource allocation and accelerate the implementation of data-based policies in waste management systems [13]. With cluster-based analysis, each region can implement strategies that are more appropriate to its characteristics and increase efficiency in the waste recycling process [14]. This study aims to develop an environmentally-based waste management optimization strategy in Indonesia using Machine Learning, by implementing K-Means Clustering in the classification of waste management units, it is expected to increase the effectiveness of the waste management system, support data-based decision making, and create a more efficient and sustainable waste management system. With Machine Learning and K-Means Clustering degradation, it can be an innovative solution in building a more adaptive, environmentally-based, and sustainability-oriented waste management system.

OBJECTIVES

Optimizing environmental-based waste management strategies is an important challenge in dealing with the impacts of urbanization and industrialization. Therefore, a more adaptive technology-based and policy-based approach is needed [15]. The risk of failure in achieving environmental targets, especially in protecting natural resources, is increasing, so a data-based risk assessment approach is a solution that needs to be implemented [16]. Studies on the spatial and temporal dynamics of the environment show that economic and demographic factors play a role in the effectiveness of waste management systems [17]. Therefore, the integration of technology-based policies, increasing public awareness, and sustainable funding are key strategies in maintaining environmental sustainability [18].

Inefficient waste management contributes to increased carbon emissions and negative impacts on ecosystems if not handled optimally. Circular economy approaches, such as recycling and waste reduction, are key solutions to reduce the volume of waste ending up in landfills [19]. In addition, food and electronic waste management play a key role in sustainability strategies, particularly in reducing the environmental impact of organic and hazardous waste [20]. Data-driven technologies are increasingly emerging in waste monitoring, where the application of machine learning can increase efficiency in analyzing waste disposal patterns [21]. With the integration of policies that support technological innovation, waste management systems can be improved to be more effective and environmentally friendly.

The role of Machine Learning in waste management is growing to improve the accuracy of waste production predictions and recycling strategies. Classification and clustering algorithms have been applied in identifying waste production patterns and anomaly detection in waste management systems [11]. Integration of big data with Machine Learning enables more accurate trend analysis, supporting data-driven decision making in waste management [14]. In addition, the combination of machine learning and sensor technology has improved automated monitoring systems, accelerated waste pattern detection, and improved waste management efficiency [8]. Therefore, Machine Learning has great potential in optimizing more sustainable and technology-based waste management strategies. In the context of waste management optimization, K-Means Clustering is an effective method in grouping regions or entities based on waste production and management characteristics. This method has been applied in various studies to optimize region segmentation based on waste production patterns and waste management efficiency [22]. However, K-Means in determining the optimal number of clusters is a challenge in its implementation, so various approaches such as Gap Statistics and distance-based methods have been developed to improve clustering accuracy [23]. Its use in industrial and urban sectors shows its effectiveness in clustering data based on waste production patterns [9]. In addition, Fuzzy K-Means is able to handle data with unclear boundaries, thus providing flexibility in clustering analysis [12].

METHODS

Method Design

This study aims to cluster provinces based on their waste production characteristics using the K-Means Clustering method. The research stages consist of data collection, data preprocessing, clustering algorithm selection, and visualization and analysis of clustering results.Pengumpulan Data

The data in this study were obtained from official sources, such as the Ministry of Environment and Forestry (KLHK) and the Central Statistics Agency (BPS). The dataset used includes several key variables related to waste production in each province, including:

a. Daily Waste Production (tons)
b. Daily Waste Volume (tons)
c. Organic Waste (tons)
d. Inorganic Waste (tons)

1. Preprocessing Data

Before clustering, the dataset is processed to ensure that the data is clean and ready to use. Preprocessing steps include:

a. Missing Value Handling: The dataset is checked for missing values (NaN) and imputed if necessary.
b. Outlier Identification: Outliers are detected using statistical methods such as Z-score or Interquartile Range (IQR).
c. Data Normalization: In order for each variable to have a uniform scale, normalization is carried out using the Min-Max Scaling or Z-score Normalization method.

2. Determining the Number of Clusters with the Elbow Method

Before implementing K-Means Clustering, the optimal number of clusters is determined using the Elbow Method. This method works by calculating the Within-Cluster Sum of Squares (WCSS) for various values of k, then plotting the results to find the elbow point that indicates the best number of clusters.

3. Implementation of K-Means Clustering

After determining the optimal number of clusters, clustering is carried out using the K-Means algorithm.

$$d(x_i, \mu_i) = \sqrt{\sum (x_i - \mu_i)^2} \tag{1}$$

4. Visualization of Clustering Results
5. Interpretation of Clustering Results

Dataset

The dataset contains information on waste production and composition in 34 provinces in Indonesia in 2023, including variables of daily waste production, volume of waste managed, organic waste, and inorganic waste, all expressed in tons. DKI Jakarta is the province with the highest waste production (2,642 tons/day), while North Kalimantan has the lowest waste production (7.52 tons/day). In general, organic waste is more dominant than inorganic waste, although there are several provinces such as South Kalimantan and Gorontalo that have a lower proportion of organic waste. In addition, there is a gap between the amount of waste produced and that which is successfully managed, indicating challenges in waste management capacity in several regions.

Table 1. Dataset Column Names

No	*Column Name*
1	No
2	Province
3	Daily Waste Production 2024(tons)

4	Daily Waste Volume 2024 (tons)
5	Organic Waste 2024 (tons)
6	Inorganic Waste 2024 (tons)

RESULTS

Clustering analysis was performed using Google Colab with the scikit-learn library, where the dataset was preprocessed to handle missing values and normalized so that each variable has a uniform scale. The Elbow method is used to determine the optimal number of clusters, in figure (1) is the result of the elbow that has been created.

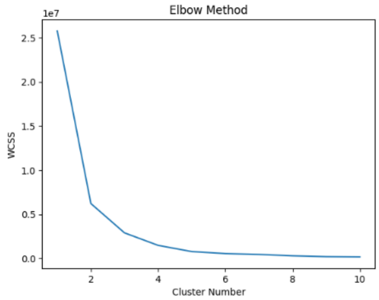

Figure 1. Elbow Chart

Figure (1) shows the Elbow Method Graph where the Within-Cluster Sum of Squares (WCSS) value decreases sharply at cluster numbers 1 to 3, then levels off after that. The elbow point seen around 3 or 4 clusters indicates the optimal number of clusters, because after this point, the decrease in WCSS is not significant. This shows that dividing the data into more than 3 or 4 clusters does not provide significant improvement in segmentation, so that the number of clusters is considered the most appropriate for this analysis. From this conclusion, the author decided to use 3 clusters.

By using 3 clusters based on the results of the Elbow Method, waste data in Indonesia was successfully grouped into three categories based on their production characteristics. Cluster 1 consists of provinces with low daily waste production, indicating areas with fewer residents or more efficient waste management systems. Cluster 2 includes provinces with medium waste production, reflecting the balance between urbanization and waste management capacity. Meanwhile, Cluster 3 contains provinces with very high waste production, which are generally metropolitan areas with high levels of consumption and economic activity. In figure (2) are the results of 3 clusters.

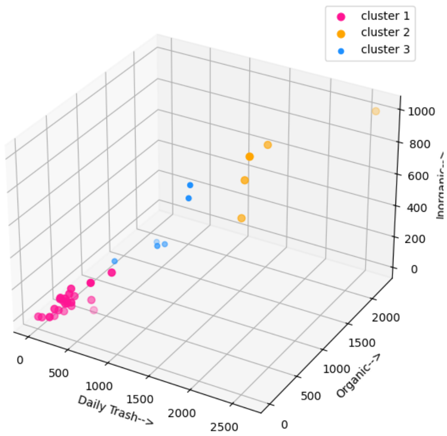

Figure 2. Cluster

In Figure (2) the clustering results with 3 clusters show that Cluster 1 (pink) consists of provinces with low waste production, Cluster 2 (orange) includes provinces with medium waste production, and Cluster 3 (blue) contains provinces with high waste production, generally metropolitan areas. The 3D scatter plot shows that the higher the daily waste production, the greater the amount of organic and inorganic waste produced. With this clustering, waste management strategies can be adjusted based on the characteristics of each group to increase the effectiveness of the waste management system.

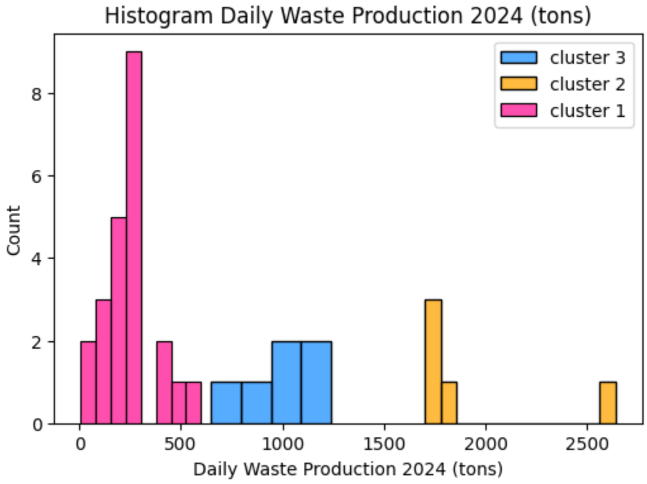

Figure 3. Characteristics of the 3 clusters

Figure (3) shows a histogram of daily waste production in 2023 showing the distribution of three main clusters. Cluster 1 (pink) includes provinces with low waste production, the majority of which are below 500 tons per day. Cluster 2 (orange) consists of provinces with medium waste production, ranging from 1,500 to 2,500 tons per day. Cluster 3 (blue) includes provinces with medium to high waste production, ranging from 500 to 1,500 tons per day. This pattern shows that most provinces have relatively low waste production, while only a few provinces produce very large amounts of waste.

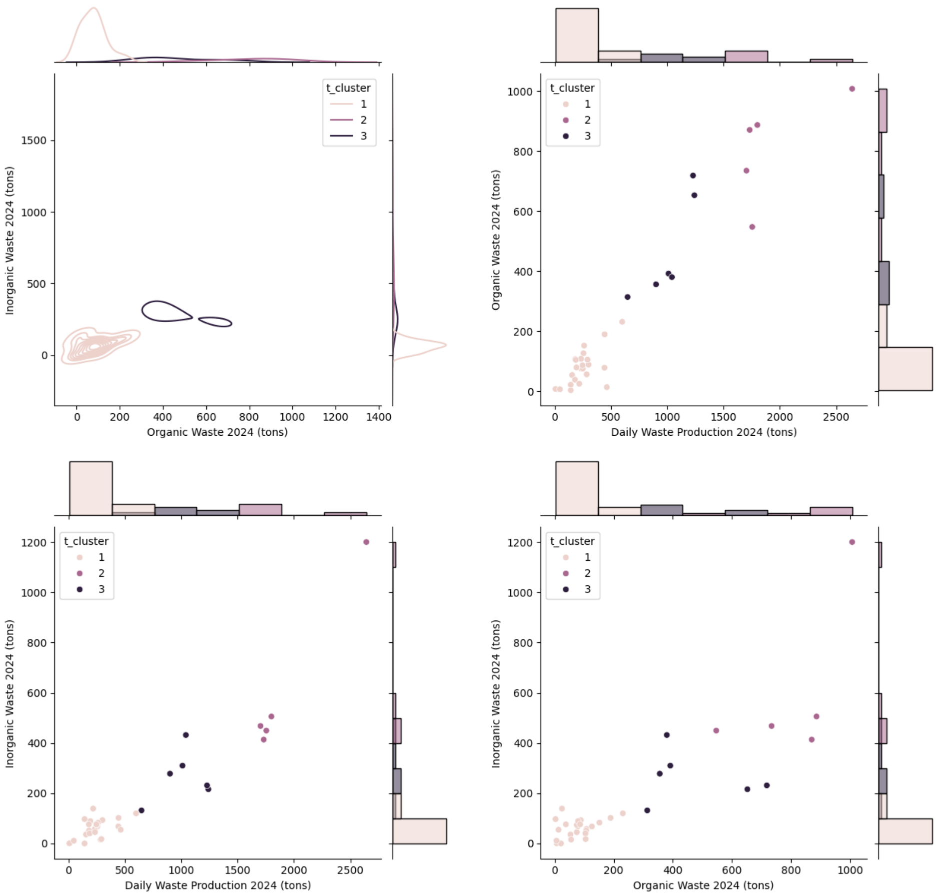

Figure 4. Join plot

Figure (4) shows the scatter plot and density plot visualization showing the relationship between organic waste, inorganic waste, and daily waste production in three clusters. Cluster 1 (light color) has a relatively small amount of organic and inorganic waste, indicating a province with low waste production. Cluster 2 (medium color) has higher waste production with a more even distribution. Cluster 3 (dark color) shows a province with very high waste production, with a significant amount of organic and inorganic waste. This graph makes it clear that the higher the daily waste production, the greater the amount of organic and inorganic waste produced.

DISCUSSION

This study shows that clustering with K-Means can help understand waste production patterns in various provinces in Indonesia. The clustering results divide the provinces into three main clusters, with the majority in the low to medium waste volume category, while only a few provinces have very high waste production. In addition, organic waste is more dominant than inorganic waste, especially in areas with high waste production. This shows that recycling, composting, and renewable energy-based management strategies can be effective solutions in reducing environmental impacts and supporting the concept of a circular economy.

The results of this clustering can be the basis for designing more appropriate waste management policies, especially in increasing waste processing capacity and encouraging community participation in reducing waste. By understanding the characteristics of each cluster, the government can develop a more adaptive data-based strategy according to the needs of its region. In addition, cluster-based mapping can also be used to support community-based recycling education and campaign programs. With the integration of data-based policies, technological innovation, and active community participation, a more efficient and sustainable waste management system can be achieved in Indonesia.

REFRENCES

[1] H. Il Kim, D. Kim, M. Mahdian, M. M. Salamattalab, S. M. Bateni, dan R. Noori, “Incorporation of water quality index models with machine learning-based techniques for real-time assessment of aquatic ecosystems,” *Environ. Pollut.*, vol. 355, no. February, hal. 124242, 2024, doi: 10.1016/j.envpol.2024.124242.

[2] C. C. S. Mello, M. M. D. Leão, dan C. C. Amorim, “Five years after the Brumadinho dam collapse: Evaluation of water quality based on combined analysis of land use and environmental data,” *Sci. Total Environ.*, vol. 957, no. October, 2024, doi: 10.1016/j.scitotenv.2024.177619.

[3] M. H. Rahaman, H. Sajjad, S. Hussain, Roshani, M. Masroor, dan A. Sharma, “Surface water quality prediction in the lower Thoubal river watershed, India: A hyper-tuned machine learning approach and DNN-based sensitivity analysis,” *J. Environ. Chem. Eng.*, vol. 12, no. 3, hal. 112915, 2024, doi: 10.1016/j.jece.2024.112915.

[4] X. Yang, W. Zhao, Y. Xu, C. D. Wang, B. Li, dan F. Nie, “Sparse K-means clustering algorithm with anchor graph regularization,” *Inf. Sci. (Ny).*, vol. 667, no. November 2023, hal. 120504, 2024, doi: 10.1016/j.ins.2024.120504.

[5] A. Bellini, B. Andersen, N. J. Klungseth, dan A. Tadayon, “Classification Machine Learning to Detect De Facto Reuse and Cyanobacteria at a Drinking,” *J. Clean. Prod.*, hal. 141753, 2024, doi: 10.1016/j.heliyon.2024.e31976.

[6] W. Wu, W. Wang, X. Jia, dan X. Feng, “Transformer Autoencoder for K-means Efficient clustering,” *Eng. Appl. Artif. Intell.*, vol. 133, no. PF, hal. 108612, 2024, doi: 10.1016/j.engappai.2024.108612.

[7] S. Schmidt dan D. Laner, “Environmental Waste Utilization score to monitor the performance of waste management systems: A novel indicator applied to case studies in Germany,” *Resour. Conserv. Recycl. Adv.*, vol. 18, no. May, hal. 200160, 2023, doi: 10.1016/j.rcradv.2023.200160.

[8] L. Jiao, X. Luo, L. Zha, H. Bao, J. Zhang, dan X. Gu, “Machine learning assisted water management strategy on a self-sustaining seawater desalination and vegetable cultivation platform,” *Comput. Electron. Agric.*, vol. 217, no. December 2023, hal. 108569, 2024, doi: 10.1016/j.compag.2023.108569.

[9] Z. Huang, H. Zheng, C. Li, dan C. Che, “Application of Machine Learning-Based K-means Clustering for Financial Fraud Detection,” *Acad. J. Sci. Technol.*, vol. 10, no. 1, hal. 33–39, 2024, doi: 10.54097/74414c90.

[10] U. Ejaz *dkk.*, “Monitoring the Industrial waste polluted stream - Integrated analytics and machine learning for water quality index assessment,” *J. Clean. Prod.*, vol. 450, no. November 2023, hal. 141877, 2024, doi: 10.1016/j.jclepro.2024.141877.

[11] S. M. Hong *dkk.*, “Using machine learning models to estimate Escherichia coli concentration in an irrigation pond from water quality and drone-based RGB imagery data,” *Water Res.*, vol. 260, no. January, hal. 121861,

2024, doi: 10.1016/j.watres.2024.121861.

[12] A. Riansyah, M. Qomaruddin, M. Indriastuti, dan M. Sagaf, "Clustering Digital Transformation of Small and Medium Enterprises (SMEs) Using Fuzzy K-means Method," *Int. Conf. Electr. Eng. Comput. Sci. Informatics*, no. September, hal. 540–544, 2023, doi: 10.1109/EECSI59885.2023.10295664.

[13] M. Sahana *dkk.*, "Assessing the effectiveness of existing early warning systems and emergency preparedness towards reducing cyclone-induced losses in the Sundarban Biosphere Region, India," *Int. J. Disaster Risk Reduct.*, vol. 90, no. March, hal. 103645, 2023, doi: 10.1016/j.ijdrr.2023.103645.

[14] C. Ortiz-Lopez, C. Bouchard, dan M. J. Rodriguez, "Ensemble machine learning using hydrometeorological information to improve modeling of quality parameter of raw water supplying treatment plants," *J. Environ. Manage.*, vol. 362, no. February, hal. 121378, 2024, doi: 10.1016/j.jenvman.2024.121378.

[15] S. Li dan T. Yang, "The impact of digital infrastructure on air quality: Mediating role of public environmental attention and industrial development," *J. Environ. Manage.*, vol. 373, no. July 2024, hal. 123893, 2025, doi: 10.1016/j.jenvman.2024.123893.

[16] Z. Nakić, D. Perković, Z. Kovač, dan B. I. Balaž, "Risk assessment of not meeting environmental objectives related to protection of human health and groundwater quality: The tiered approach in the context of the EU water framework directive," *Groundw. Sustain. Dev.*, vol. 29, no. October 2024, 2025, doi: 10.1016/j.gsd.2025.101428.

[17] Z. Dong, X. Luo, K. Zhao, J. Ge, dan I. Y. S. Chan, "Comprehensive assessment method for building environmental performance: Trade-off between indoor environmental quality and life cycle carbon emissions," *Build. Environ.*, vol. 272, no. August 2024, hal. 112598, 2025, doi: 10.1016/j.buildenv.2025.112598.

[18] A. W. G. C. N. Wijethunga, M. M. Rahman, dan T. Sarker, "Analyzing financial development's moderating role on the investment-environmental quality nexus: Evidence from Australia using Autoregressive Distributed Lag approach," *Environ. Sustain. Indic.*, vol. 25, no. September 2024, hal. 100568, 2025, doi: 10.1016/j.indic.2024.100568.

[19] Y. N. Kavil, H. Alshemmari, M. M. Alkasbi, S. S. Alelyani, M. I. Orif, dan R. K. Al-Farawati, "Electronic waste and environmental diplomacy: How GCC E-waste management interfaces with the Stockholm Convention," *J. Hazard. Mater. Adv.*, vol. 17, no. September 2024, hal. 100610, 2025, doi: 10.1016/j.hazadv.2025.100610.

[20] M. J. Muñoz-Torres, I. Ferrero-Ferrero, J. V. Gisbert-Navarro, dan J. M. Rivera-Lirio, "Environmental assessment of food loss and waste prevention and reduction solutions: Navigating the complexity of integrating stakeholders' decisions," *Environ. Impact Assess. Rev.*, vol. 112, no. August 2024, 2025, doi: 10.1016/j.eiar.2024.107788.

[21] F. Bahmei, N. Bahramifar, S. Ghasemi, H. Younesi, dan M. Weil, "Jo ur l P re r f," *J. Clean. Prod.*, hal. 145145, 2025, doi: 10.1016/j.jclepro.2025.145145.

[22] I. K. Khan, H. Daud, N. Zainuddin, dan R. Sokkalingam, "Standardizing reference data in gap statistic for selection optimal number of cluster in K-means algorithm," *Alexandria Eng. J.*, vol. 118, no. February 2024, hal. 246–260, 2025, doi: 10.1016/j.aej.2025.01.034.

[23] H. S. Fouda, S. I. Farghaly, dan H. S. Dawood, "Weighted joint LRTs for cooperative spectrum sensing using K-means clustering," *Phys. Commun.*, vol. 67, no. October, hal. 102528, 2024, doi: 10.1016/j.phycom.2024.102528.