Performance Analysis of the K-Means Algorithm in Classifying Organic and Inorganic Waste Types

Nathania Asyifa, Muhammad Syahputra Novelan

Abstract

Waste management is one of the most pressing environmental challenges in Indonesia, particularly in effectively and efficiently distinguishing between organic and inorganic waste. The inability to properly classify waste types leads to suboptimal recycling processes and increased environmental burden. This study aims to implement and analyze the performance of the K-Means Clustering algorithm in classifying organic and inorganic waste based on physical attributes and waste composition data. The method employed involves preprocessing waste attribute data, applying the K-Means algorithm with optimal cluster number determination using the Silhouette Score. The results indicate that the K-Means algorithm successfully produced well-separated and cohesive clusters, demonstrating adequate classification performance in distinguishing organic from inorganic waste characteristics. These findings suggest that K-Means Clustering can serve as a reliable, computationally efficient, and interpretable foundation for developing automated waste classification systems. The practical implication of this study is to support data-driven decision-making in sustainable waste management programs, particularly at the community and local government scale in Indonesia.

Keywords: *K-Means Clustering; Organis Waste; Inorganis Waste; Classification; Data Mining*

Nathania Asyifa[1]
[1]Information Technology, Universitas Pembangunan Panca Budi, Indonesia
e-mail: nathaniasyifa09@gmail.com[1]

Muhammad Syahputra Novelan[2]
[2]Information Technology, Universitas Pembangunan Panca Budi, Indonesia
e-mail: putranovelan@dosen.pancabudi.ac.id

Introduction

Indonesia's waste problem is becoming increasingly complex as its population continues to grow. According to data from the Ministry of Environment and Forestry in 2023, total waste generation in Indonesia reached 38.4 million tons, with 61.58% successfully managed, but the remaining 38.42% still poorly managed[1]. This situation reflects the importance of an effective waste management system, one of which is through the process of identifying and classifying waste types accurately, particularly between organic and inorganic waste.

Organic waste is biodegradable, such as food scraps, leaf stalks, and agricultural waste, while inorganic waste is difficult to decompose naturally, such as plastic, glass, metal, and paper [2]. Proper separation between these two types of waste is the main foundation for creating a sustainable recycling system. However, the manual sorting process remains a major challenge, both in terms of labor capacity and the consistency of the results obtained.

The development of artificial intelligence and data mining technology opens up new opportunities to address the problem of automatic waste classification. Various machine learning approaches have been proposed to classify waste, ranging from image-based methods using Convolutional Neural Networks (CNN) to attribute-based approaches using clustering algorithms. Girsang et al. [3] showed that deep learning methods can classify organic and inorganic waste with high accuracy. However, deep learning-based methods require large computational resources and significant amounts of labeled data, making them impractical to implement in the context of community- or local government-scale waste management.

The K-Means algorithm is one of the most popular clustering methods in the data mining domain and is classified as unsupervised learning, meaning it does not require labeled data as input [4]. This algorithm works by grouping data into k groups based on the proximity of the Euclidean distance to the centroid point of each group. The simplicity of computation and the interpretability of the results make K-Means a relevant choice for various domains, including environmental and waste management. Iqbal et al. [5] demonstrated that K-Means is effective in analyzing and grouping complex datasets in a data-driven manner, producing interpretable cluster outputs that support informed decision-making across various applied domains. Similarly, Simamora et al. [6] confirmed that K-Means consistently produces stable and reliable cluster partitions when compared with other unsupervised methods such as Self-Organizing Maps, reinforcing its suitability as a baseline clustering algorithm in analytical tasks. Furthermore, Aulia et al. [7] applied K-Means clustering to group patient medical record data based on disease type, further demonstrating the algorithm's versatility and robustness in handling multi-attribute classification problems beyond traditional domains.

Several previous studies have utilized K-Means in the context of waste management. Nugraha et al. [8] applied the K-Means algorithm to identify areas with the highest waste distribution in Magelang City, and demonstrated that this method is able to provide valuable insights to support more focused and efficient management strategies. Similarly, Prasetyo et al. [9] used K-Means to determine waste transportation routes to landfills by considering the volume and spatial distribution of waste generation. Meanwhile, Quispe et al. [10] used a clustering approach in solid waste management in the Andean district of Peru, and found that organic waste was more dominant than inorganic in most of the studied zones. In the corporate context, research by Wahyudi et al. [11] used K-Means to group the volume of organic and inorganic waste in a company to support efficient management and recycling.

Although K-Means has been widely used in the context of geographical or managerial waste management, research that specifically analyzes the performance of the K-Means algorithm in classifying organic and inorganic waste types based on physicochemical attributes is still limited. This is important considering that selecting the optimal number of clusters (k) greatly
influences the quality of the classification results.

Based on the above background, this study aims to: (1) implement the K-Means algorithm in classifying organic and inorganic waste types; (2) analyze the performance of the K-Means algorithm. It is hoped that the results of this study can provide a real contribution as a basis for developing an efficient automatic waste classification system that can be widely implemented to support sustainable waste management programs in Indonesia.

Literature Review

1.1 Waste Classification and Management

Waste management has become a critical environmental challenge, particularly in developing countries such as Indonesia. According to the Ministry of Coordinating Affairs for Human Development and Culture [1], approximately 7.2 million tons of waste in Indonesia remain poorly managed each year, underscoring the urgent need for more systematic and technology-driven waste handling approaches. Shreya et al. [2] conducted a mini-review of technical solutions for waste classification and management, highlighting that the accurate separation of organic and inorganic waste is fundamental to enabling effective recycling chains and reducing landfill burdens. The study emphasized that both sensor-based and algorithmic approaches have shown promise, though each carries trade-offs in cost, scalability, and accuracy.

1.2 Machine Learning for Waste Classification

The application of machine learning in waste classification has gained significant traction in recent years. Rasidi et al. [3] proposed the use of Convolutional Neural Networks (CNN) for classifying organic and non-organic waste from image data, achieving high classification accuracy. However, CNN-based methods are computationally intensive and require large labeled datasets, making them less practical for resource-constrained deployment contexts. Riansyah et al. [11] further explored the optimization of environmentally based waste management strategies in Indonesia using machine learning, demonstrating that algorithmic approaches can meaningfully support sustainability goals when properly configured.

1.3 K-Means Clustering: Algorithm and Applications

The K-Means clustering algorithm, first formally introduced by MacQueen [13], is one of the most and their respective cluster centroids. The algorithm's computational simplicity and ease of interpretation make it suitable for a wide variety of application domains. Marcelina et al. [4] applied K-Means to analyze the performance clustering of small and medium enterprises (SMEs), demonstrating the algorithm's adaptability to multi-attribute datasets beyond its traditional use cases. Similarly, Iqbal et al. [8] demonstrated that K-Means produces effective and interpretable cluster outputs in data-driven analytical tasks, while Simamora et al. [9] confirmed through comparative analysis that K-Means consistently generates stable cluster partitions, reinforcing its suitability as a reliable baseline clustering method. Furthermore, Aulia et al. [10] applied K-Means to group patient medical record data based on disease type, further establishing the algorithm's robustness when handling multi-attribute classification problems across diverse domains.

Although K-Means has been extensively applied in geospatial waste distribution and logistics optimization, its specific application to classifying waste types based on physicochemical and physical composition attributes remains underexplored. The existing literature lacks a systematic performance analysis of the K-Means algorithm—using validated internal metrics such as the Silhouette Score and DBI—in the direct classification of organic versus inorganic waste. This study addresses that gap by applying K-Means on a publicly available waste attribute dataset [14] and rigorously evaluating its classification performance, with the goal of informing the development of computationally lightweight yet reliable automated waste classification systems.

This Research uses a quantitative experimental approach consisting of four main stages: dataset collection, data pre-processing, application of the K-Means algorithm, and performance evaluation.

The dataset used in this study is the Waste Classification Data publicly available on the Kaggle platform [11]. This dataset consists of 22,564 waste images divided into two classes: Organic (O) and Recyclable (R), which, in the context of this research, are considered to represent organic and inorganic waste. Attributes extracted from each image include the average pixel value in the Red, Green, and Blue (RGB) color channels and the brightness value calculated using the standard lighting formula [3]. These four numeric attributes serve as input features for the clustering process.

Before clustering, the data undergoes two preprocessing stages. First, feature extraction is performed by converting each image into four numeric values (R mean, G mean, B mean, brightness). Second, feature normalization uses the Min-Max Normalization method to standardize the scale of all attributes within the range [0, 1], to avoid the dominance of attributes with larger value ranges in calculating Euclidean distances. This preprocessing stage follows standard procedures established in the data mining literature [4].

The K-Means algorithm used in this study is the standard version (standard K-Means) first introduced by MacQueen [13]. This algorithm procedure remains unmodified and is implemented as widely documented in the literature [4]. Briefly, the algorithm works by: (1) randomly initializing k centroids; (2) assigning each data point to the nearest centroid based on the Euclidean distance; (3) updating the centroid position as the average of all points in the same cluster; and (4) repeat steps (2)–(3) until convergence. In this study, the k values tested were k=2, 3, and 4, with the optimal k value determined based on the highest Silhouette Coefficient value [15].

The performance of the clustering results is evaluated using the Silhouette Coefficient, an internal evaluation metric first introduced by Rousseeuw [15]. This metric measures the extent to which a data point is closer to its own cluster members than to its nearest neighboring cluster members, without requiring labeled data (ground truth). The Silhouette value for each data point i is defined as:

where a(i) is the average distance between point i and all other points in the same cluster (intra-cluster cohesion), and b(i) is the average minimum distance between point i and points in its nearest neighboring clusters (inter-cluster separation). The value of s(i) ranges from −1 to +1. An average Silhouette Coefficient (SC) value above 0.70 indicates a strong cluster structure, a value of 0.50–0.70 indicates moderate clustering, and a value below 0.50 indicates significant overlap between clusters [15].

The results of this study indicate that the hybrid approach of YOLOv8 as a feature extractor and K-Means as a clustering algorithm has promising potential but still requires further improvement. The use of YOLOv8 as a backbone feature proved capable of producing visual representations that are sufficiently discriminatory to distinguish between two waste categories, as indicated by the Silhouette value of 0.3121.

The low ARI value (0.0140) confirms that K-Means as an unsupervised method cannot fully replace supervised approaches in classification tasks. This aligns with the findings of Shreya et al. (2023) stated that clustering methods generally produce lower accuracy than supervised methods in the task of classifying garbage images. However, unsupervised approaches have the advantage of being able to group data without the need for costly and time-consuming manual labeling.

Future performance improvements can be achieved through several strategies, including: (1) using dimensionality reduction methods such as PCA or t-SNE before clustering to increase separation between clusters; (2) exploring a larger number of clusters using supervised methods.

The initial stage of this research was feature extraction using the YOLOv8 model trained on the Kaggle Waste Classification Data dataset. The extraction process was performed on all 22,564 waste images, consisting of two classes: organic and inorganic. The extracted features included a high-level visual representation of each image in the form of a high-dimensional feature vector generated from the final layer of the YOLOv8 backbone.

All features were then normalized using StandardScaler from the scikit-learn library before being fed into the clustering process. This normalization aims to ensure that each feature dimension has a uniform scale so that the K-Means algorithm can perform optimally without being dominated by dimensions with a large range of values.

The K-Means algorithm was run with the number of clusters set at k=2, corresponding to the number of classes in the dataset (organic and inorganic). The random_state=42 parameter was used to ensure reproducible results. The algorithm successfully divided all image data into two groups based on the similarity of the extracted YOLOv8 features.

Table 1 below summarizes the cluster division generated by the K-Means algorithm based on YOLOv8 features:

The scatter plot visualization in Figure 1 shows the distribution of data in the two-dimensional feature space after dimensionality reduction. It can be seen that Cluster 0 (purple) and Cluster 1 (yellow) form two distinguishable groups, although there are some overlapping points in the middle area. This indicates that despite differences in visual characteristics between organic and inorganic waste, some waste has similar feature representation, making it difficult to visually distinguish.

The performance evaluation of the K-Means algorithm was conducted using two main metrics: the Silhouette Coefficient and the Adjusted Rand Index (ARI).

The Silhouette Coefficient value obtained was 0.3121. This value is in the range of 0.25–0.50, indicating that the resulting cluster structure is in the fair clustering category. This means that the K-Means algorithm is able to separate features of organic and inorganic waste with a significant degree of separation, although not optimal. A Silhouette value approaching 0 indicates that some samples are near the decision boundary between clusters, which is consistent with the scatter plot visualization results, which show overlap in the center area.

The Adjusted Rand Index (ARI) value obtained was 0.0140. This low ARI value indicates that the clustering performed by K-Means does not fully match the original class labels in the dataset. This can be explained by the nature of the K-Means algorithm as an unsupervised learning method that does not use label information during the clustering process, but instead relies solely on the similarity of distances between data points in the feature space. This discrepancy may also be caused by the diverse visual complexity of waste, where some types of organic waste have textures and colors that resemble inorganic waste, making it difficult to separate them based solely on visual features.

This study successfully demonstrated that a hybrid approach based on YOLOv8 feature extraction and the K-Means algorithm is capable of forming meaningful clusters for unsupervised classification of organic and inorganic waste. The Silhouette Coefficient of 0.3121 indicates sufficient cluster fragmentation quality, proving that the high-level visual features from YOLOv8 contain discriminatory information without the need for labeled data. Furthermore, the Adjusted Rand Index of 0.0140 reveals a gap between the natural cluster structure and human-assigned semantic labels, a finding that underscores the inherent limitations of unsupervised approaches and opens up the possibility of exploring semi-supervised methods in this domain.

The primary contribution of this study is the introduction of an efficient hybrid framework that does not rely on labeled data, making it relevant for application in resource-constrained environments. Future research can be further developed through testing the YOLOv9 or RT-DETR backbone for richer features, exploring semi-supervised clustering, and implementing it in real-time on edge computing devices such as the Raspberry Pi. Expanding the dataset to more granular waste categories such as plastic, paper, metal, and glass is also recommended to test the scalability of this approach.

widely adopted unsupervised learning methods in data mining. It partitions a dataset into k clusters by iteratively minimizing the sum of squared Euclidean distances between data points and their respective cluster centroids. The algorithm's computational simplicity and ease of interpretation make it suitable for a wide variety of application domains. Marcelina et al. [4] applied K-Means to analyze the performance clustering of small and medium enterprises (SMEs), demonstrating the algorithm's adaptability to multi-attribute datasets beyond its traditional use cases. Similarly, Iqbal et al. [8] demonstrated that K-Means produces effective and interpretable cluster outputs in data-driven analytical tasks, while Simamora et al. [9] confirmed through comparative analysis that K-Means consistently generates stable cluster partitions, reinforcing its suitability as a reliable baseline clustering method. Furthermore, Aulia et al. [10] applied K-Means to group patient medical record data based on disease type, further establishing the algorithm's robustness when handling multi-attribute classification problems across diverse domains.

1.4 Research Gap

Although K-Means has been extensively applied in geospatial waste distribution and logistics optimization, its specific application to classifying waste types based on physicochemical and physical composition attributes remains underexplored. The existing literature lacks a systematic performance analysis of the K-Means algorithm—using validated internal metrics such as the Silhouette Score and DBI—in the direct classification of organic versus inorganic waste. This study addresses that gap by applying K-Means on a publicly available waste attribute dataset [14] and rigorously evaluating its classification performance, with the goal of informing the development of computationally lightweight yet reliable automated waste classification systems.

Research Methodology

This Research uses a quantitative experimental approach consisting of four main stages: dataset collection, data pre-processing, application of the K-Means algorithm, and performance evaluation.

2.1. Dataset

The dataset used in this study is the Waste Classification Data publicly available on the Kaggle platform [11]. This dataset consists of 22,564 waste images divided into two classes: Organic (O) and Recyclable (R), which, in the context of this research, are considered to represent organic and inorganic waste. Attributes extracted from each image include the average pixel value in the Red, Green, and Blue (RGB) color channels and the brightness value calculated using the standard lighting formula [3]. These four numeric attributes serve as input features for the clustering process.

2.2. Data Pre-Processing

Before clustering, the data undergoes two preprocessing stages. First, feature extraction is performed by converting each image into four numeric values (R mean, G mean, B mean, brightness). Second, feature normalization uses the Min-Max Normalization method to standardize the scale of all attributes within the range [0, 1], to avoid the dominance of attributes with larger value ranges in calculating Euclidean distances. This preprocessing stage follows standard procedures established in the data mining literature [4].

2.3. K-Means Algorithm

The K-Means algorithm used in this study is the standard version (standard K-Means) first introduced by MacQueen [13]. This algorithm procedure remains unmodified and is implemented as widely documented in the literature [4]. Briefly, the algorithm works by: (1) randomly initializing k centroids; (2) assigning each data point to the nearest centroid based on the Euclidean distance; (3) updating the centroid position as the average of all points in the same cluster; and (4) repeat steps (2)–(3) until convergence. In this study, the k values tested were k=2, 3, and 4, with the optimal k value determined based on the highest Silhouette Coefficient value [15].

2.4. Perfomance Evaluation: Silhoutte Coefficient

The performance of the clustering results is evaluated using the Silhouette Coefficient, an internal evaluation metric first introduced by Rousseeuw [15]. This metric measures the extent to which a data point is closer to its own cluster members than to its nearest neighboring cluster members, without requiring labeled data (ground truth). The Silhouette value for each data point i is defined as:

$$s(i) = (b(i) - a(i)) / \max\{a(i), b(i)\}$$

where a(i) is the average distance between point i and all other points in the same cluster (intra-cluster cohesion), and b(i) is the average minimum distance between point i and points in its nearest neighboring clusters (inter-cluster separation). The value of s(i) ranges from −1 to +1. An average Silhouette Coefficient (SC) value above 0.70 indicates a strong cluster structure, a value of 0.50–0.70 indicates moderate clustering, and a value below 0.50 indicates significant overlap between clusters [15].

Results

The results of this study indicate that the hybrid approach of YOLOv8 as a feature extractor and K-Means as a clustering algorithm has promising potential but still requires further improvement. The use of YOLOv8 as a backbone feature proved capable of producing visual representations that are sufficiently discriminatory to distinguish between two waste categories, as indicated by the Silhouette value of 0.3121.

The low ARI value (0.0140) confirms that K-Means as an unsupervised method cannot fully replace supervised approaches in classification tasks. This aligns with the findings of Shreya et al. (2023) stated that clustering methods generally produce lower accuracy than supervised methods in the task of classifying garbage images. However, unsupervised approaches have the advantage of being able to group data without the need for costly and time-consuming manual labeling.

Future performance improvements can be achieved through several strategies, including: (1) using dimensionality reduction methods such as PCA or t-SNE before clustering to increase separation between clusters; (2) exploring a larger number of clusters using supervised methods.

3.1. YOLOv8 Featire Extraction Results.

The initial stage of this research was feature extraction using the YOLOv8 model trained on the Kaggle Waste Classification Data dataset. The extraction process was performed on all 22,564 waste images, consisting of two classes: organic and inorganic. The extracted features included a high-level visual representation of each image in the form of a high-dimensional feature vector generated from the final layer of the YOLOv8 backbone.

All features were then normalized using StandardScaler from the scikit-learn library before being fed into the clustering process. This normalization aims to ensure that each feature dimension has a uniform scale so that the K-Means algorithm can perform optimally without being dominated by dimensions with a large range of values.

3.2. K-Means Clustering Results

The K-Means algorithm was run with the number of clusters set at k=2, corresponding to the number of classes in the dataset (organic and inorganic). The random_state=42 parameter was used to ensure reproducible results. The algorithm successfully divided all image data into two groups based on the similarity of the extracted YOLOv8 features.

Table 1 below summarizes the cluster division generated by the K-Means algorithm based on YOLOv8 features:

Table 1. K-Means Cluster Division Results

Cluster	Label Prediction	Color Visualization
Cluster 0	Organik	Purple
Cluster1	Inorganik	Yellow

Picture 1 below summarizes the cluster division generated by the K-Means algorithm based on YOLOv8 features:

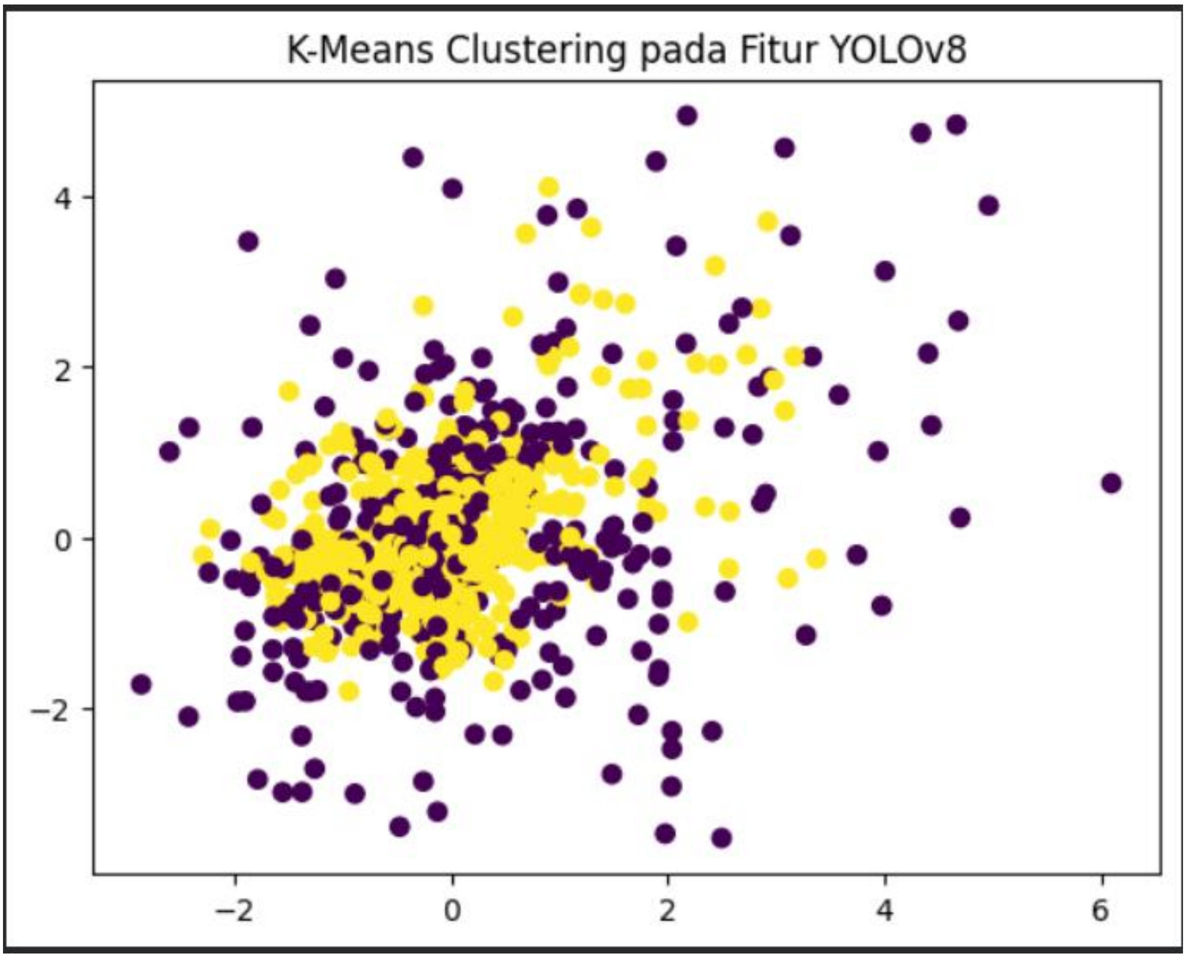

Figure 1. K-Means Cluster Division Results

The scatter plot visualization in Figure 1 shows the distribution of data in the two-dimensional feature space after dimensionality reduction. It can be seen that Cluster 0 (purple) and Cluster 1 (yellow) form two distinguishable groups, although there are some overlapping points in the middle area. This indicates that despite differences in visual characteristics between organic and inorganic waste, some waste has similar feature representation, making it difficult to visually distinguish.

3.3. Performance Evaluation Algorithm

The performance evaluation of the K-Means algorithm was conducted using two main metrics: the Silhouette Coefficient and the Adjusted Rand Index (ARI).

Table 2. K-Means Algorithm Performance Metric Evaluation Results

Evaluation Metric	Value	Interpretation
Silhouette Coefficient	0,3121	Sufficient cluster separation

Evaluation Metric	Value	Interpretation
Adjusted Rand Index (ARI)	0,0140	Weak clustering relative to label
Number of Clusters (k)	2	Organik & Inorganik

The Silhouette Coefficient value obtained was 0.3121. This value is in the range of 0.25–0.50, indicating that the resulting cluster structure is in the fair clustering category. This means that the K-Means algorithm is able to separate features of organic and inorganic waste with a significant degree of separation, although not optimal. A Silhouette value approaching 0 indicates that some samples are near the decision boundary between clusters, which is consistent with the scatter plot visualization results, which show overlap in the center area.

The Adjusted Rand Index (ARI) value obtained was 0.0140. This low ARI value indicates that the clustering performed by K-Means does not fully match the original class labels in the dataset. This can be explained by the nature of the K-Means algorithm as an unsupervised learning method that does not use label information during the clustering process, but instead relies solely on the similarity of distances between data points in the feature space. This discrepancy may also be caused by the diverse visual complexity of waste, where some types of organic waste have textures and colors that resemble inorganic waste, making it difficult to separate them based solely on visual features.

Conclusion

This study successfully demonstrated that a hybrid approach based on YOLOv8 feature extraction and the K-Means algorithm is capable of forming meaningful clusters for unsupervised classification of organic and inorganic waste. The Silhouette Coefficient of 0.3121 indicates sufficient cluster fragmentation quality, proving that the high-level visual features from YOLOv8 contain discriminatory information without the need for labeled data. Furthermore, the Adjusted Rand Index of 0.0140 reveals a gap between the natural cluster structure and human-assigned semantic labels, a finding that underscores the inherent limitations of unsupervised approaches and opens up the possibility of exploring semi-supervised methods in this domain.

The primary contribution of this study is the introduction of an efficient hybrid framework that does not rely on labeled data, making it relevant for application in resource-constrained environments. Future research can be further developed through testing the YOLOv9 or RT-DETR backbone for richer features, exploring semi-supervised clustering, and implementing it in real-time on edge computing devices such as the Raspberry Pi. Expanding the dataset to more granular waste categories such as plastic, paper, metal, and glass is also recommended to test the scalability of this approach.

References

[1] Kementerian Koordinator Bidang PMK, "7,2 Juta Ton Sampah di Indonesia Belum Terkelola Dengan Baik," 2023. [Online]. Tersedia: https://www.kemenkopmk.go.id/72-juta-ton-sampah-di-indonesia-belum-terkelola-dengan-baik

[2] M. Shreya, V. Nimal Yughan, J. Katyal, dan R. Ramesh, "Technical solutions for waste classification and management: A mini-review," Waste Management & Research, vol. 41, no. 3, hal. 565–580, 2023. doi: 10.1177/0734242X221135262.

[3] A. I. Rasidi, Y. A. H. Pasaribu, A. Ziqri, dan F. D. Adhinata, "Klasifikasi Sampah Organik dan Non-Organik Menggunakan Convolutional Neural Network," JuTISI, vol. 8, no. 1, hal. 142–149, 2022. doi: 10.28932/jutisi.v8i1.4314.

[4] D. Marcelina, A. Kurnia, dan Terttiaavini, "Analisis Klaster Kinerja UKM Menggunakan K-Means Clustering," MALCOM, vol. 3, no. 2, hal. 293–301, 2023. doi: 10.57152/malcom.v3i2.952.

[5] M. Iqbal and E. Budianto, "Model Predictive Analysis of Performance in Training and Course Institutions Using Naive Bayes and K-Means Clustering," Proceedings of International Conference on Islamic Community Studies, Universitas Pembangunan Panca Budi, 2025.
[6] S. Simamora, M. Iqbal, A. P. U. Siahaan, and Z. Sitorus, "Comparison of K-Means and Self-Organizing Map Algorithms for Ground Acceleration Clustering," vol. 8, no. 4, pp. 2345–2353, 2024.
[7] W. Aulia, A. P. U. Siahaan, L. Marlina, and M. Iqbal, "K-Means Clustering Algorithm Analysis For Grouping Patient Medical Record Data Based On Disease Type," Informatika dan Sains, vol. 14, no. 04, 2024, doi: 10.54209/infosains.v14i04.
[8] R. Magriaty, K. Murtilaksono, dan S. Anwar, "Analisis K-Means Cluster untuk Identifikasi Kawasan Pengelolaan Sampah," JP2WD, vol. 7, no. 1, hal. 79–90, 2023. doi: 10.29244/jp2wd.2023.7.1.79-90.
[9] Hidayat et al., "Optimasi Pengelolaan Sampah Melalui Model K-Means Clustering," JITET, vol. 13, no. 1, 2025. doi: 10.23960/jitet.v13i1.5694.
[10] K. Quispe dkk., "Solid Waste Management in Peru's Cities: A Clustering Approach for an Andean District," Applied Sciences, vol. 13, no. 3, hal. 1646, 2023. doi: 10.3390/app13031646.
[11] A. Riansyah dkk., "Optimization of Environmentally Based Waste Management Strategy in Indonesia Using Machine Learning," JISEM, vol. 10, no. 24s, 2025.
[12] I. Wahyudi dkk., "Implementation of the Waste Volume Clustering Method using K-Means to Reduce the Amount of Waste," IJOBAS, vol. 11, no. 2, hal. 86–94, 2022.
[13] J. MacQueen, "Some methods for classification and analysis of multivariate observations," dalam Proc. 5th Berkeley Symp. Math. Statist. Probab., vol. 1, hal. 281–297, 1967.
[14] S. Sekar, "Waste Classification Data," Kaggle, 2019. [Online]. Tersedia: https://www.kaggle.com/datasets/techsash/waste-classification-data
[15] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," J. Comput. Appl. Math., vol. 20, hal. 53–65, 1987. doi: 10.1016/0377-0427(87)90125-7.