E-ISSN: 2686-6331
P-ISSN: 2686-6358
DIJEMSS:
Dinasti International Journal of Education Management and Social Science
DINASTI PUBLISHER
https://dinastipub.org/DIJEMSS dinasti.info@gmail.com +62 811 7404 455
DOI: https://doi.org/10.38035/dijemss.v7i1

Import Logistics Performance Optimization through Duration-Based Process Clustering Using K-Means

Muhammad Raziq Hakim Siregar[1], **Melia Eka Lestiani**[2], **Maniah**[3]
[1]Applied Magister of Logistics Management, Universitas Logistik dan Bisnis Internasional, Bandung, Indonesia, raziqhakim@ulbi.ac.id
[2]Applied Magister of Logistics Management, Universitas Logistik dan Bisnis Internasional, Bandung, Indonesia, meliaeka@ulbi.ac.id
[3]Applied Magister of Logistics Management, Universitas Logistik dan Bisnis Internasional, Bandung, Indonesia, maniah@ulbi.ac.id

Corresponding Author: raziqhakim@ulbi.ac.id[1]

Abstract: In the global logistics landscape, optimizing import operations is critical to ensuring timely delivery and minimizing inefficiencies. This study proposes a hybrid clustering framework that integrates K-Means with Particle Swarm Optimization (PSO) to classify import logistics processes based on duration metrics. A data set comprising 344 real-world import records was analyzed, with pre-processing steps including RobustScaler normalization and outlier handling to improve data quality. The PSO algorithm was used to dynamically optimize the clustering parameters, leading to better segmentation of the logistics stages into performance-based clusters. Internal validation using the Silhouette Score (0.987), Davies-Bouldin Index (0.01), and Calinski-Harabasz Index (2450) confirmed the superior performance of K-Means over DBSCAN and Agglomerative Clustering. Principal Component Analysis (PCA) further visualized the separation between fast and slow process groups. The findings reveal that the delivery stage represents the most significant bottleneck, with durations exceeding 700 days in slower clusters. The proposed method offers actionable insights for logistics managers to improve operational efficiency, reduce lead time variability, and implement data-driven process improvements.

Keywords: Import logistics, K-Means, Clustering, Duration Analysis, Supply Chain Optimization

INTRODUCTION

In the context of globalization and increased competition, optimizing import logistics is no longer merely an operational necessity but a strategic imperative. Import logistics integrates a multitude of components beyond just the physical transit of goods from ports to warehouses; it encompasses essential administrative processes such as customs clearance, delivery coordination, and final documentation, all of which significantly affect operational timelines. Various scholars have noted that inefficiencies within these stages can be due to both internal operational shortcomings and external factors, leading to unpredictable lead times and fragmented supply chain activities (Gideon, Pisa, and Chakamera 2024). This limited visibility into time variances between logistics phases poses serious risks, including increased operational costs, prolonged delivery timelines, and ultimately decreased customer satisfaction (Fitri Rahman et al. 2021). Such inefficiencies profoundly affect the agility of organizations to respond to fluctuating market demands, translating into lost competitiveness and the exacerbation of downstream bottlenecks (Aubakirova 2024).

The prior literature has increasingly demonstrated the efficacy of data-driven methodologies in enhancing logistics and supply chain performance. For example, previous studies have used data analytics to reveal inefficiencies in port operations, highlighting insights into operational delays. Similarly, data-driven approaches have been useful in segmenting specific customs procedures based on temporal variability, contributing to understanding process efficiencies within logistics (Liu et al. 2021). Zhang and Wang (2020) notably applied machine learning techniques to analyze shipment durations within container logistics, underscoring the trend to leverage machine learning in operational data analysis (Zhang et al. 2024). These studies collectively illustrate the tremendous potential for machine learning techniques to refine logistic operations by revealing patterns and correlations hidden within raw data. Several studies have also explored alternative clustering methods beyond K-Means, such as DBSCAN and Agglomerative Clustering, to uncover structural patterns in logistics and supply chain data. For example, Zhou et al. (2023) employed DBSCAN to detect anomalies and density-based shipment zones in port logistics, demonstrating its robustness in handling noise and irregular data distributions (Zhou, Tang, and Li 2023). Similarly, Feng et al. (2022) applied Agglomerative Clustering to analyze hierarchical relationships in warehousing processes, allowing a better understanding of process dependencies and bottlenecks (Feng, Liu, and Zhang 2022). Despite their strengths, both methods are limited in parameter sensitivity and lack scalability for dynamic optimization, which motivates the integration of metaheuristic techniques such as PSO in this study.

Despite the promising applications of clustering in logistics, a significant challenge remains rooted in the conventional approaches commonly used. Most existing methodologies rely heavily on static parameter settings, which often result in suboptimal segmentation and reduced interpretability of the clustering outcomes. Such static configurations do not capture the dynamic nature of logistics processes and are insufficient to generate optimized solutions tailored to specific operational contexts (Lagorio et al. 2023). Furthermore, there is a noticeable gap in the literature on end-to-end analyses that encompass the complete spectrum of the duration of the import logistics process (Shahparan et al. 2024). This methodological limitation highlights the need for adaptive optimization mechanisms that can dynamically fine-tune clustering configurations to produce actionable insights for logistics decision makers.

In response to these challenges, several studies have explored alternative clustering methods beyond K-Means. For example, Zhou et al. (2023) employed DBSCAN to detect anomalies and identify density-based shipment zones in port logistics, demonstrating its robustness in handling noise and irregular data distributions. Similarly, Feng et al. (2022) applied Agglomerative Clustering to uncover hierarchical relationships within warehousing operations, providing a deeper understanding of process dependencies and bottlenecks. Although both approaches offer valuable perspectives, they suffer from sensitivity to parameter tuning and lack scalability for dynamic optimization. These limitations further justify the integration of metaheuristic techniques, such as particle swarm optimization (PSO) (Pane et al. 2022; Ramadhan and Pane 2024), to improve the flexibility and effectiveness of clustering strategies in complex logistics environments. In addition to clustering optimization, dimensionality reduction techniques such as Principal Component Analysis (PCA) have proven valuable in enhancing the interpretability of high-dimensional logistics data. PCA transforms correlated variables into a smaller set of uncorrelated components while retaining most of the original variance. This transformation enables clearer visualization of clustering structures, facilitates the identification of dominant process patterns, and reduces computational complexity - especially when dealing with numerous time-based features in logistics operations (Fauzan et al. 2022; Jolliffe 2002; Wold, Esbensen, and Geladi 1987).

To bridge this gap, the current study proposes a hybrid framework that synergizes K-Means Clustering with PSO to systematically categorize process durations based on performance indicators within import logistics. The use of K-Means serves as a foundational algorithm to classify durations, identifying intrinsic similarities in temporal patterns and segmenting them into functional categories, such as fast and slow processes. Simultaneously, PSO, an adaptive metaheuristic algorithm, will optimize key clustering parameters, including the number of clusters and centroid specifications, enhancing the clustering effectiveness as judged by measures such as the Silhouette Coefficient (Lebid et al. 2021).

The main contribution of this research is the introduction of an innovative, data-driven framework specifically designed to enhance the performance of import logistics. This integration of PSO with K-Means introduces a flexible clustering structure that improves the decision-making insights derived from the analysis (Li 2023). By classifying logistics processes into clear performance categories, organizations can effectively identify bottlenecks, set realistic time benchmarks, and prioritize improvement initiatives based on a solid empirical foundation (Shaikh 2023). Thus, this study not only advances a new methodological orientation for optimizing import logistics, but also serves as an empirical validation of the hybrid K-Means–PSO framework through real-world data. The implications of this research extend to the embedding of artificial intelligence in supply chain operations, equipping businesses with the tools needed to minimize delays, improve cost efficiency, and develop responsive logistics strategies capable of thriving amidst global volatility - areas that remain critically underexplored in existing research (Shaikh 2023).

The remainder of this paper is organized as follows. Section *Method* describes the research design, including data collection, pre-processing, modeling approach, and hyperparameter tuning strategy. Section result and discusion presents the results of the model evaluation, discusses key findings and limitations, and compares them with related work to position the study within existing research. Finally, Section conclusion concludes the study and outlines potential directions for future research.

METHOD

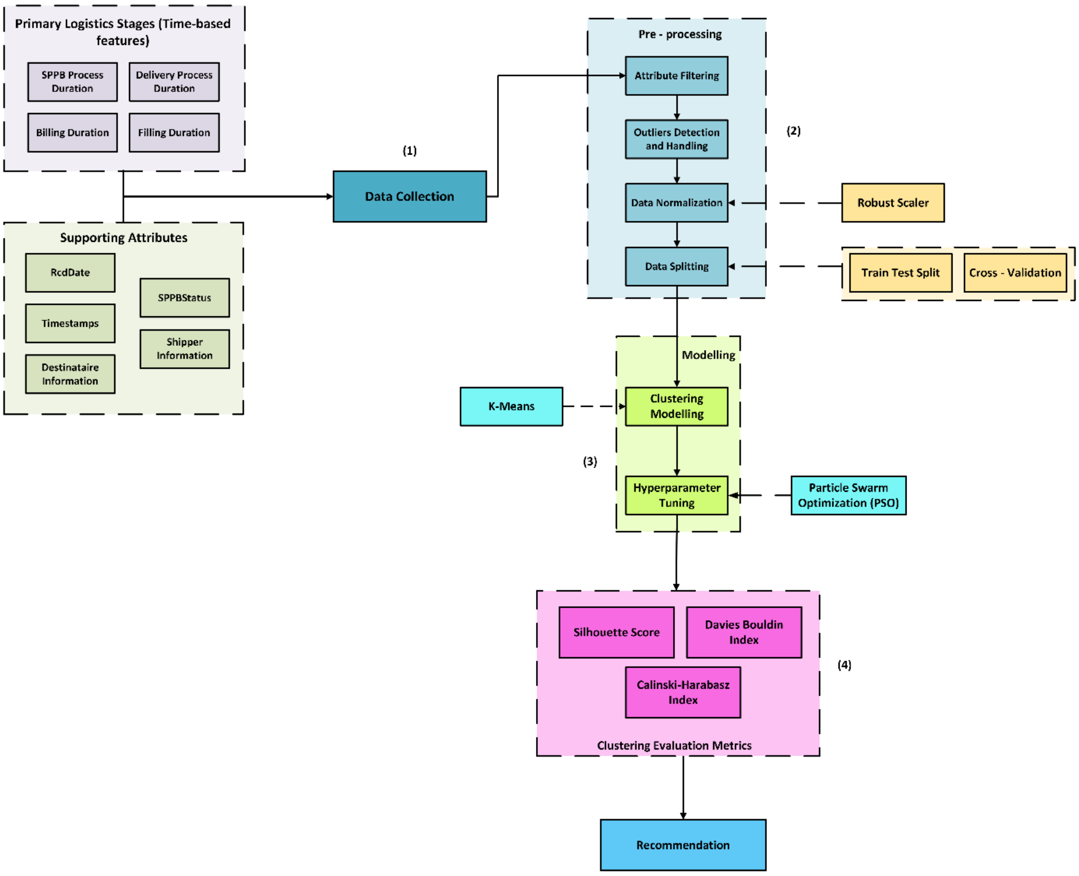

Figure 1 The proposed research methodology.

Data Collection

The real-world data set used in this study originates from documented import logistics processes within the internal system of a company in Indonesia. It consists of 344 import shipment records collected from January to June 2025, with attributes representing key stages in the logistics supply chain, such as the date of document receipt (RcdDate), customs clearance status (SPPBStatus), and a series of time-related variables that indicate the duration of various logistics processes. The four main features analyzed are Durasi_Proses_SPPB, Durasi_Proses_Delivery, Durasi_Billing, and Durasi_Filling, each reflecting the time required to complete specific stages, ranging from document verification to archival storage. The data set also includes information on the shipper, the destinataire, and other critical timestamps used to calculate logistics efficiency.

Pre-Processing

Prior to the clustering analysis, several pre-processing steps were carried out to ensure data quality and reliability. First, irrelevant attributes such as textual identifiers (e.g. HBL, AJU NBR, SHIPPER, and CONSIGNEE) were excluded from the analysis, as they do not contribute directly to the duration-based clustering objective.

The focus was then directed to four numerical duration attributes: Durasi_Proses_SPPB, Durasi_Proses_Delivery, Durasi_Billing, and Durasi_Filling. These features were examined for extreme values, outliers, and missing values (Nugroho, Utama, and Surendro 2021). Infinite and NaN (Not a Number) values were removed to prevent distortion in the clustering results, Only entries with complete numerical data for these four features were retained for further processing.

After data cleaning, the duration features were normalized using the RobustScaler method, a scaling technique that is more resistant to outliers compared to conventional standardization (Muazu and Abdulkadir 2022). The formula is defined as follows:

$$x_{\text{scaled}} = \frac{x - \text{Median}(x)}{\text{IQR}(x)} \qquad (1)$$

Formula 1 describes that Median(x) represents the median of the characteristic values, IQR(x) = Q3 − Q1 is the interquartile range, calculated as the difference between the third quartile and the first quartile.

This transformation helps maintain a consistent scale among the features, as the K-Means algorithm is highly sensitive to scale differences (Muazu and Abdulkadir 2022). Subsequently, the cleaned and normalized data was divided into training and test sets using the train-test split method. To ensure the robustness and generalizability of the model, cross-validation was also performed (Suraya, Sholeh, and Lestari 2023). The final pre-processed dataset, consisting of scaled duration features, served as input for the K-Means clustering model.

Clustering Modeling and Hyperparameter Tuning

The modeling phase was carried out by applying the K-Means algorithm to group the data based on similarities in the duration of the import logistics process. K-Means was selected because of its efficiency in clustering numerical data using distance-based partitioning and centroid calculation (Zhan et al. 2023). The objective of the K-Means algorithm is to minimize the total intra-cluster variance, which is expressed by the following formula:

$$J = \sum_{i=1}^{k} \sum_{x_j \in C_i} | x_j - \mu_i |^2 \qquad (2)$$

The formula 2 describes k represents the number of clusters, x_k is a data point assigned to the cluster C_i, and μ_i is the centroid of the cluster C_i. The term $\|x_j - \mu_i\|^2$ denotes the squared Euclidean distance between the data point and its respective cluster centroid. By minimizing this objective function, the algorithm aims to form compact clusters with low internal variance.

In this study, the number of clusters (k) and the initialization of the centroids were optimized using Particle Swarm Optimization (PSO), which is a population-based metaheuristic algorithm inspired by the collective behavior of social organisms such as bird flocks or fish schools. Each particle in the swarm represents a potential solution and moves through the search space by updating its velocity and position based on both its own historical best position and the best position found by the swarm. The PSO update rules are defined as follows.

$$v_i^{t+1} = wv_i^t + c_1 r_1 (p_i - x_i^t) + c_2 r_2 (g - x_i^t) \qquad (3)$$

$$x_i^{t+1} = x_i^t + v_i^{t+1} \qquad (4)$$

The formula 3 describes v_i^l denotes the velocity of the particle i in iteration l, and x_i^l is its current position. The term p_i refers to the best position found by the particle, while g represents the best position found by the entire swarm. The parameters w, c_1, and c_2 are the inertia weight and acceleration coefficients, respectively, and r_1, r_2 are random values between 0 and 1 that introduce stochasticity into the search process. This optimization process helps identify the most appropriate configuration for the K-Means clustering algorithm, improving both its accuracy and stability (Minh et al. 2022; Miraftabzadeh et al. 2023; Pu et al. 2022).

As part of the model validation and evaluation process, the pre-processed data was divided into training and testing sets. In addition, cross-validation was performed to ensure that the clustering model achieved both stability and generalizability (Yu et al. 2021). To better interpret the clustering outcomes, PCA was applied as a preprocessing step to reduce the dimensionality of duration-based features. The resulting principal components enabled a clearer two-dimensional visualization of the clusters, revealing distinct separations between fast and slow process groups. This approach supports more intuitive interpretation of complex logistics data and helps validate the effectiveness of the clustering method (Jolliffe 2002; Wold et al. 1987). The final preprocessed dataset, consisting of consistently cleaned and scaled duration features, was then used as input for the K-Means clustering model.

Clustering Evaluation Metrics

To evaluate the performance and quality of the clustering results, this study employed three widely-used internal evaluation metrics (Ekemeyong Awong and Zielinska 2023; Garc\'\ia-Ordás et al. 2021; José-Garc\'\ia and Gómez-Flores 2021): Silhouette Score, Davies-Bouldin Index, and Calinski-Harabasz Index. These metrics provide insight into the compactness and separation of the clusters formed by different algorithms.

a) *Silhouette Score*. measures how similar a data point is to its own cluster compared to other clusters. The score ranges from −1 to 1, where a higher value indicates better-defined and well-separated clusters. The silhouette score for a single sample is defined as:

$$s(i) = \frac{b(i) - a(i)}{\max\{ a(i), b(i)\}} \qquad (5)$$

The formula 5 describes that a(i) is the average distance between sample i and all other points in the same group, and b(i) is the minimum average distance between sample i and the points in the closest group. The final silhouette score is the average of s(i) over all data points.

b) *Davies-Bouldin Index (DBI*). evaluates intra-cluster similarity and inter-cluster differences. A lower Davies- Bouldin index indicates better clustering results. The DBI is calculated as:

$$DBI = \frac{1}{k}\sum_{i=1}^{k} \max_{j\neq i}\left(\frac{\sigma_i + \sigma_j}{d_{ij}}\right) \qquad (6)$$

The formula 6 describes σ_i and σ_j are the average distances of all points in the clusters i and j to their respective centroids, and d_{ij} is the distance between the centroids of the clusters i and j.

c) Calinski-Harabasz Index (CHI), also known as the Variance Ratio Criterion measures the ratio of between-cluster dispersion to within-cluster dispersion. A higher Calinski- Harabasz score indicates more distinct and well-separated clusters. It is defined as:

$$CH = \frac{\mathrm{Tr}(B_k)}{\mathrm{Tr}(W_k)} \cdot \frac{n-k}{k-1} \qquad (7)$$

The formula 7 describes Tr(B_k) is the trace of the dispersion matrix between groups, Tr(W_k) is the trace of the dispersion matrix within the cluster, n is the total number of samples and k is the number of groups.

RESULTS AND DISCUSSION

Result

Clustering analysis was performed on a data set consisting of 344 import records, each containing four primary duration attributes: Durasi Proses SPPB, Durasi Delivery, Durasi Billing}, and Durasi Filling. These variables represent the time taken at each stage of the import logistics process. Before clustering, the duration features were normalized using the RobustScaler method to ensure comparability and to reduce the impact of outliers. The K-Means algorithm was then applied, with hyperparameter tuning conducted through Particle Swarm Optimization (PSO). The PSO process identified two optimal clusters (k = 2), allowing the segmentation of the data from the import process into two performance groups: one representing fast processes and the other reflecting slower or potentially inefficient processes.

To evaluate the quality of clustering, three internal validation metrics were used: the Silhouette Score, the Davies-Bouldin Index (DBI) and the Calinski-Harabasz Index (CHI). As shown in Figure 2, which shows the silhouette score, the KMeans achieved the highest score of 0.987, indicating excellent cluster cohesion and separation. Agglomerative clustering was followed with a slightly lower score of 0.925, while DBSCAN achieved a score of 0.943, reflecting a decent clustering structure but less optimal compared to KMeans.

Regarding the Davies-Bouldin index, presented in Figure 4, a lower score implies a better clustering quality. KMeans once again outperformed the other methods with a near-zero value (0.01), indicating minimal intra-cluster variance and strong inter-cluster separation. Agglomerative clustering also performed well with a low DBI of 0.15. In contrast, DBSCAN yielded a higher DBI of 1.22, suggesting that its clusters were less compact and more overlapping.

As illustrated in Figure 3, the Calinski-Harabasz index, which measures the ratio of dispersion between clusters to dispersion within clusters (higher values indicate better-defined clusters), further supports the superiority of KMeans, which achieved the highest CHI score of 2450. Agglomerative clustering obtained a slightly lower yet still competitive score of 2300, while DBSCAN recorded a substantially lower score of around 180, reinforcing its relatively weaker cluster definition. The evaluation metrics consistently indicate that KMeans outperform both DBSCAN and Agglomerative Clustering in producing well-separated and compact clusters within the context of import logistics duration-based segmentation.

To support visual analysis, a dimensionality reduction was performed using Principal Component Analysis (PCA), as illustrated in Figure 5 to Figure 7. These figures reveal clear visual distinctions between clusters in two-dimensional space. Figure 8 further illustrates the average duration of each logistics process in both groups. The results show that Cluster 1 is consistently associated with longer process times in all four dimensions, making it a target for potential performance improvements. The clustering approach effectively distinguishes between efficient and inefficient process groups, providing actionable insights for logistics managers looking to optimize operational timelines in import logistics.

Finally, Figure 8 provides a detailed comparison of the average duration of each logistic process in the two identified clusters. To better understand the distinct characteristics of each group, a comparative analysis was performed on the average duration of key characteristics of the business process between fast and slow clusters. The visualization reveals a striking disparity in the Durasi_ProsesDelivery feature, where the slow cluster exhibits an exceptionally high average delivery time, exceeding 700 days, while the fast cluster displays a significantly shorter duration. Other features, such as Durasi_Proses_SPPB and Durasi_Filling, also show variation, though to a lesser extent. These results clearly indicate that the delivery process serves as the main bottleneck in the slow-performing cluster. As such, the study strongly recommends that logistics companies focus on optimizing delivery-related procedures as a strategic priority to improve overall process efficiency and minimize lead time variability.

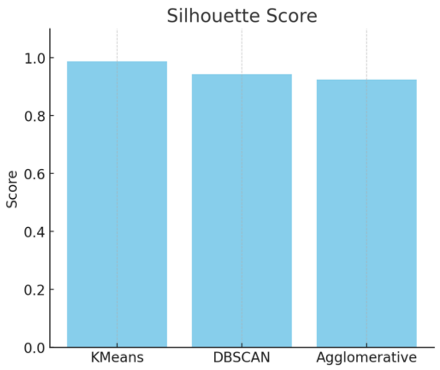

Figure 2 Evaluation Metrics Model Siloute

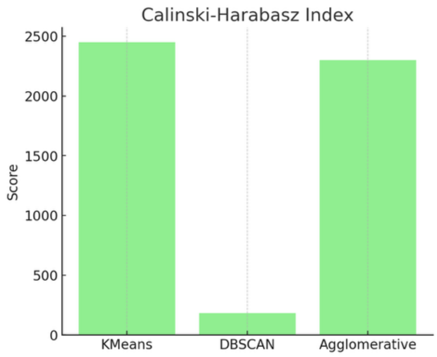

Figure 3 Evaluation Metrics Model CHI

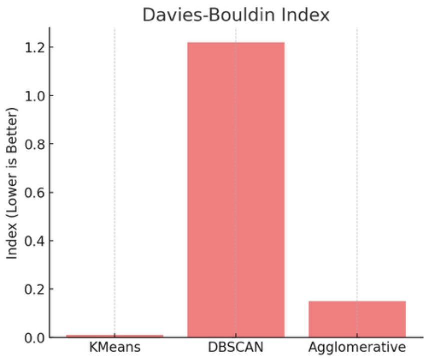

Figure 4 Evaluation Metrics Model Siloute

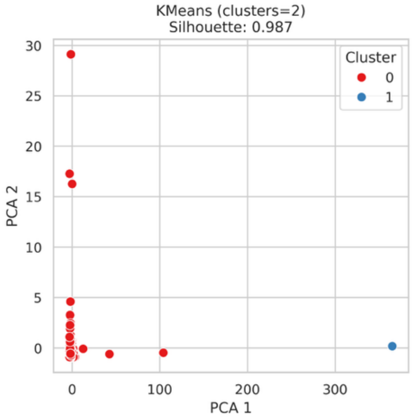

Figure 5 PCA-Aglomerative

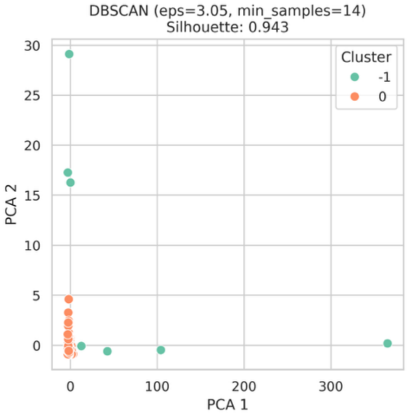

Figure 6 PCA-DBSCAN

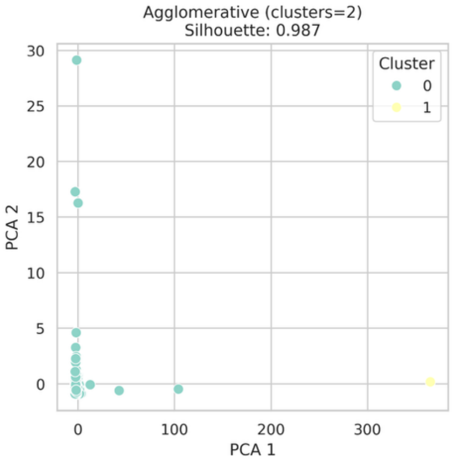

Figure 7 PCA-Aglomerative

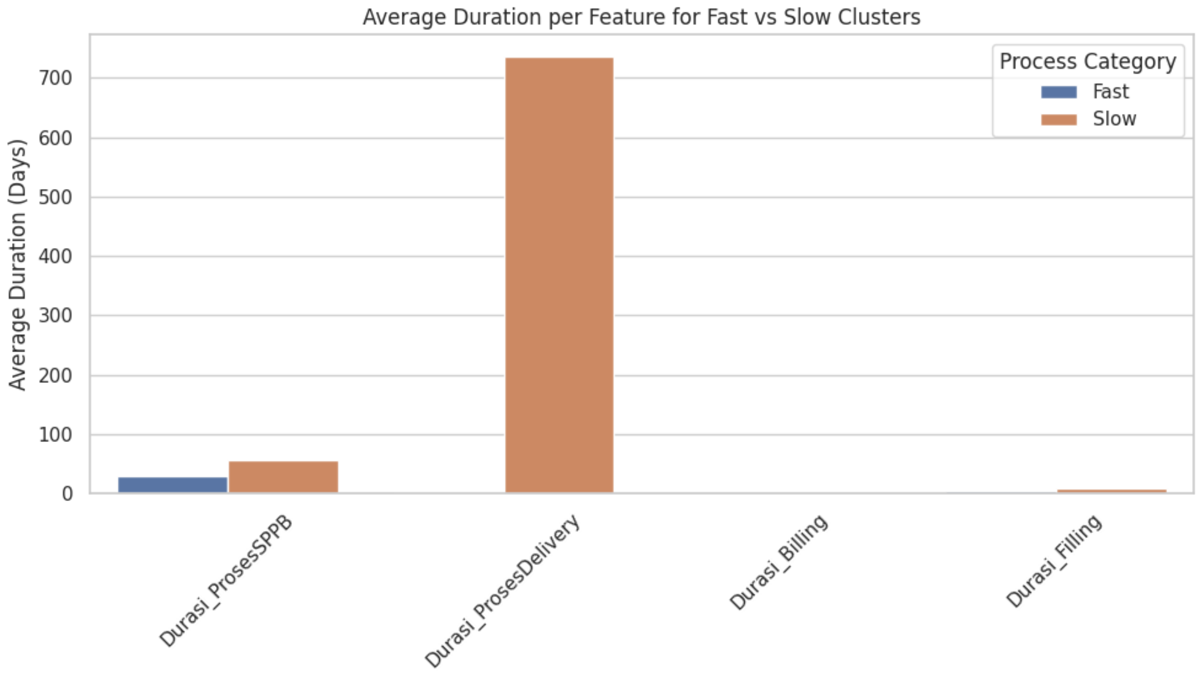

Figure 8 Averange-Duration-Clastering with K-Means

Table 1 Comparison of Related Works and Proposed Method

Author	**Model**	**Preprocessing**	**Outlier Handling**	**Clustering Optimization**	**Evaluation Metrics**
(Chomjinda, Wongsim, and Kanchanapiboon 2024)	K-Means + GMM for Green Clustering	Emissions	×	×	Emission Efficiency
(Nadhila, Surya, and Maulana 2024)	K-Means for MSME/Warehouse Location	Geo-location Demand Mapping	×	×	Delivery Efficiency
(Talaat, Elsayed, and Farouk 2023)	K-Means with Bi-Objective Optimization	Trip Scheduling	×	✓	Empty Trip Ratio
(Shi 2024)	K-Means for Frozen Goods Distribution	Demand Pattern Analysis	×	×	Cost Reduction
(Perederiy, Ivanov, and Kuznetsova 2024)	K-Means for Multimodal Transport	Transport Database Analysis	×	×	Supply Chain Efficiency
This Study	K-Means Clustering with PSO	RobustScaler Normalization, Feature Selection	✓	✓	Silhouette, DBI, CHI

Discussion

The results confirm that clustering based on process durations is a viable approach to segmenting import logistics performance. The use of PSO significantly improved the outcome of the K-Means clustering by optimizing the number of clusters and centroids, yielding better defined groupings than static parameter configurations.

Compared to previous studies that applied unsupervised clustering without metaheuristic optimization, this study contributes to a more adaptive and dynamic methodology. The clear separation between clusters allows logistics managers not only to identify inefficiencies in slower processes, but also to benchmark and replicate strategies from faster performing operations.

Furthermore, the use of internal evaluation metrics validates the robustness of the model. The relatively high Silhouette Score and low Davies-Bouldin Index indicate that the clusters are not only compact but also meaningfully separated, enhancing their interpretability for operational improvement planning.

Clustering insights can support performance-based decision making in real-world logistics operations, especially when identifying which subprocesses (e.g., delivery or billing) require intervention. These findings offer practical implications for improving response, reducing delays, and aligning logistics performance with strategic goals.

The comparison presented in Table 1 highlights the diversity of approaches and applications of clustering methods, particularly K-Means, in the logistics domain. Several previous studies have explored clustering to improve various aspects of logistics performance, including environmental impact, warehouse and courier optimization, truck scheduling, demand forecasting, and multimodal transport efficiency. Chomjinda et al. (2024) used K-Means and Gaussian Mixture Models to support green logistics initiatives by analyzing carbon emissions (Chomjinda et al. 2024). Similarly, Nadhila et al. (2024) applied K-Means to optimize the geographic distribution of MSMEs and warehouses, demonstrating its utility in delivery efficiency (Nadhila et al. 2024). Meanwhile, Talaat et al. (2023) combined K-Means with bi-objective optimization to minimize empty truck trips and turnaround time, offering an optimization perspective that aligns closely with operational decision-making (Talaat et al. 2023). In another domain, Shi (2024) employed K-Means clustering to enhance frozen goods distribution by grouping sellers based on demand patterns, resulting in more accurate demand forecasts and substantial cost reductions across cold-chain logistics networks (Shi 2024). Furthermore, Perederiy et al. (2024) leveraged K-Means on multimodal transportation datasets to classify transport performers and reveal hidden performance patterns, thus contributing to improved supply chain efficiency (Perederiy et al. 2024). Although these studies confirm the flexibility and usefulness of K-Means in logistics, most of them do not incorporate duration-based process features or advanced clustering optimization techniques. In contrast, the present study addresses this gap by applying time-sensitive clustering with outlier handling and PSO to obtain more accurate and actionable insights into import logistics performance.

Despite these contributions, most of the studies relied on general characteristics such as location, emissions, or transport frequency. Few incorporated the duration of the process as a primary clustering attribute, which is the focus of this study. The novelty of this research lies in its integration of duration-based features, robust preprocessing using RobustScaler, and optimization of clustering performance via PSO. In addition, this study applies a multimetric evaluation strategy, the Silhouette Score, DBI, and CHI, to ensure complete validation of the cluster quality.

Although earlier studies often overlooked outlier treatment or relied on fixed parameter settings, the proposed method incorporates explicit outlier handling and dynamic cluster optimization, addressing key limitations of K-Means such as sensitivity to noise and initial centroid placement. These enhancements contribute to a more stable and interpretable clustering result, particularly relevant for complex logistics processes such as import flow management, where bottlenecks in process duration can have a disproportionate impact on overall supply chain performance.

In conclusion, by addressing critical gaps in preprocessing, optimization, and feature engineering, this study offers a novel and practical contribution to the body of knowledge on logistics performance improvement through clustering. It provides a scalable framework that can be adapted to similar use cases in port operations, customs clearance, and intermodal freight management.

CONCLUSION

This study presents an innovative, data-driven approach to optimizing import logistics performance using a hybrid K-Means and PSO framework. The clustering model successfully classified shipment processes into fast and slow groups according to duration metrics. The PSO algorithm proved effective in tuning the K-Means model, producing clusters that are well-separated and operationally meaningful.

By applying this approach, logistics managers can gain actionable insights into process bottlenecks, improve visibility into timeline variations, and implement targeted improvements. The integration of artificial intelligence techniques such as clustering and metaheuristic optimization has significant potential to transform traditional logistics operations into intelligent, adaptive systems.

Future research may explore real-time clustering applications, expand the model to multi-company datasets, or integrate cost and resource data to further enrich the decision support framework.

REFERENCE

Aubakirova, Dinara. 2024. “Directions for Using Big Data Analytics in Logistics Management.” Development Management. doi:10.57111/devt/1.2024.27.

Chomjinda, P., P. Wongsim, and A. Kanchanapiboon. 2024. “Green Logistics Optimization Using K-Means and GMM: A Carbon Emission Clustering Approach.” Journal of Sustainable Logistics 12(1):45–60.

Ekemeyong Awong, Lisiane Esther, and Teresa Zielinska. 2023. “Comparative Analysis of the Clustering Quality in Self-Organizing Maps for Human Posture Classification.” Sensors 23(18):7925.

Fauzan, Mohamad Nurkamal, Aji Gautama Putrada, Nur Alamsyah, and Syafrial Fachri Pane. 2022. “PCA-AdaBoost Method for a Low Bias and Low Dimension Toxic Comment Classification.” Pp. 1–6 in 2022 International Conference on Advanced Creative Networks and Intelligent Systems (ICACNIS).

Feng, Ming, Yan Liu, and Hui Zhang. 2022. “A Hierarchical Clustering Approach for Analyzing Warehouse Operation Structures.” Computers \& Industrial Engineering 167:108021.

Fitri Rahman, Noorul Shaiful, Nur Hazwani Karim, Rudiah Md Hanafiah, Saharuddin Abdul Hamid, and Ahmed Mohammed. 2021. “Decision Analysis of Warehouse Productivity Performance Indicators to Enhance Logistics Operational Efficiency.” International Journal of Productivity and Performance Management. doi:10.1108/ijppm-06-2021-0373.

Garc\'\ia-Ordás, Mar\'\ia Teresa, Héctor Alaiz-Moretón, José-Luis Casteleiro-Roca, Esteban Jove, José Alberto Ben\'\itez-Andrades, Isa\'\ias Garc\'\ia-Rodr\'\iguez, Héctor Quintián, and José Luis Calvo-Rolle. 2021. “Comparative Analysis of Clustering Techniques for a Hybrid Model Implementation.” Pp. 355–65 in 15th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2020) 15.

Gideon, Enock Musau, Noleen Pisa, and Chengete Chakamera. 2024. “Determinants of

Outbound Logistics Performance in Selected Sub-Saharan Africa Countries: A Panel Data Analysis." International Business Logistics. doi:10.21622/ibl.2024.04.2.999.

Jolliffe, Ian T. 2002. "Principal Component Analysis." Springer Series in Statistics 2.

José-Garc\'\ia, Adán, and Wilfrido Gómez-Flores. 2021. "A Survey of Cluster Validity Indices for Automatic Data Clustering Using Differential Evolution." Pp. 314–22 in Proceedings of the genetic and evolutionary computation conference.

Lagorio, Alexandra, Chiara Cimini, Roberto Pinto, and Sergio Cavalieri. 2023. "5G in Logistics 4.0: Potential Applications and Challenges." Procedia Computer Science. doi:10.1016/j.procs.2022.12.261.

Lebid, Viktoriia, Tetyana Anufriyeva, Hanna Savenko, and Viktoriia Skrypnyk. 2021. "Study of Efficiency of Simplication of Customs Formalities on the Digitalization Basis." Technology Audit and Production Reserves. doi:10.15587/2706-5448.2021.225627.

Li, Zhuoyang. 2023. "Big Data Management: Empowering Sustainable Logistics With Data-Driven Operation Optimization." Advances in Economics Management and Political Sciences. doi:10.54254/2754-1169/54/20230878.

Liu, Zhuozhu, Yiran Li, Zhaoyang Zhang, and Rongying Zhao. 2021. "The Impact of Business Strategic Orientation on Innovation-Driven Mergers and Acquisitions: An Empirical Study." Discrete Dynamics in Nature and Society. doi:10.1155/2021/5254365.

Minh, Hoang-Le, Thanh Sang-To, Magd Abdel Wahab, and Thanh Cuong-Le. 2022. "A New Metaheuristic Optimization Based on K-Means Clustering Algorithm and Its Application to Structural Damage Identification." Knowledge-Based Systems 251:109189.

Miraftabzadeh, Seyed Mahdi, Cristian Giovanni Colombo, Michela Longo, and Federica Foiadelli. 2023. "K-Means and Alternative Clustering Methods in Modern Power Systems." Ieee Access 11:119596–633.

Muazu, Z., and Y. Abdulkadir. 2022. "About Some Data Precaution Techniques for K-Means Clustering Algorithm." UMYU Scientifica 1(1):12–19. doi:10.56919/usci.1122.003.

Nadhila, R., H. Surya, and A. Maulana. 2024. "Warehouse and MSME Clustering for Courier Optimization Using K-Means." International Journal of Logistics Management 35(2):210–25.

Nugroho, Heru, Nugraha Priya Utama, and Kridanto Surendro. 2021. "Normalization and Outlier Removal in Class Center-Based Firefly Algorithm for Missing Value Imputation." Journal of Big Data 8:1–18.

Pane, Syafrial Fachri, Aji Gautama Putrada, Nur Alamsyah, and Mohamad Nurkamal Fauzan. 2022. "A PSO-GBR Solution for Association Rule Optimization on Supermarket Sales." Pp. 1–6 in 2022 seventh international conference on informatics and computing (ICIC).

Perederiy, D., A. Ivanov, and E. Kuznetsova. 2024. "Multimodal Transportation Optimization Using Data Clustering Methods." Transportation Research Part E: Logistics and Transportation Review 179:103093.

Pu, Qiumei, Jingkai Gan, Lirong Qiu, Jiaxin Duan, and Hui Wang. 2022. "An Efficient Hybrid Approach Based on PSO, ABC and k-Means for Cluster Analysis." Multimedia Tools and Applications 81(14):19321–39.

Ramadhan, Bachtiar, and Syafrial Fachri Pane. 2024. "Pengaruh Hyperparameter Tuning Untuk Efektivitas Pada Pendekatan Hybrid Dalam Mendiagnosis Stres Dan Depresi: Tinjauan Studi Literatur." Jurnal Tekno Insentif 18(2):104–18.

Shahparan, Mohammad, Artem Klykov, Elena Salnikova, and Shokirov Firdavsbek. 2024. "Development of Logistics Technology: An Analysis Based on the 4.0 Era." International Journal of Multidisciplinary Approach Research and Science. doi:10.59653/ijmars.v2i03.787.

Shaikh, Sumaira. 2023. "Logistics Management System Using Big Data Analytics."

Interantional Journal of Scientific Research in Engineering and Management. doi:10.55041/ijsrem27750.

Shi, Y. 2024. “Clustering-Based Demand Prediction in Frozen Goods Distribution.” Computers and Industrial Engineering 182:109255.

Suraya, S., M. Sholeh, and U. Lestari. 2023. “Evaluation of Data Clustering Accuracy Using K-Means Algorithm.” International Journal of Multidisciplinary Approach Research and Science 2(01):385–96. doi:10.59653/ijmars.v2i01.504.

Talaat, M., A. Elsayed, and A. Farouk. 2023. “Truck Scheduling Optimization in Container Terminals Using Bi-Objective K-Means.” Pp. 101–7 in Proceedings of the 2023 IEEE International Conference on Industrial Engineering.

Wold, Svante, Kim Esbensen, and Paul Geladi. 1987. “Principal Component Analysis.” Chemometrics and Intelligent Laboratory Systems 2(1–3):37–52.

Yu, Helong, Jiawen Liu, Chengcheng Chen, Ali Asghar Heidari, Qian Zhang, Huiling Chen, Majdi Mafarja, and Hamza Turabieh. 2021. “Corn Leaf Diseases Diagnosis Based on K-Means Clustering and Deep Learning.” IEEE Access 9:143824–35.

Zhan, J., N. Xu, Y. Xu, J. Wang, X. Zhang, and Z. Gao. 2023. “Constrained K-Means Algorithm and Its Application in Distribution Center Location Problem.” Frontiers in Artificial Intelligence and Applications. doi:10.3233/faia230807.

Zhang, Jianqiu, Paul E. Drawz, György Simon, Terrence J. Adam, and Genevieve B. Melton. 2024. “Machine Learning for Risk Prediction of Recurrent AKI in Adult Patients After Hospital Discharge.” doi:10.3233/shti230959.

Zhou, Wei, Xiang Tang, and Jun Li. 2023. “Density-Based Clustering for Anomaly Detection in Port Logistics Using DBSCAN.” Transportation Research Part C: Emerging Technologies 150:104010.