BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 6, October 2026 | Hal 2639-2646
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i6.1345

Analysis of Food Distribution Efficiency Clustering Based on Cost and Delivery Time Using the K-Means Algorithm

Muhammad Syahputra Novelan*, Solly Aryza, Puspita Wanny

Pascasarjana, Program Studi Teknologi Informasi, Universitas Pembangunan Panca Budi, Medan, Indonesia
Email: [1,*]putranovelan@dosen.pancabudi.ac.id, [2]sollyaryzalubis@gmail.com, [3]puspitawanny142@gmail.com
Correspondence Author Email: putranovelan@dosen.pancabudi.ac.id
Received: **16/08/2026**, Revised: **27/10/2026**, Accepted: **30/10/2026**, Published**: 31/10/2026**

Abstract−Food distribution efficiency is an important factor in maintaining supply chain performance, particularly in controlling transportation costs and ensuring timely delivery. Inefficiencies in distribution may lead to increased operational expenses and delays in delivering food products to their destinations. This study aims to analyze food distribution efficiency by grouping distribution activities based on transportation costs and delivery time using the K-Means clustering algorithm. The research stages consist of data collection, data preprocessing, feature selection, data standardization, determination of the optimal number of clusters, implementation of the K-Means algorithm, and cluster evaluation. Transportation cost and delivery time are used as the main variables to identify distribution patterns and classify distribution activities into efficiency groups. The resulting clusters are intended to distinguish distribution activities with relatively efficient, moderate, and less efficient characteristics, providing a basis for identifying operational patterns that require improvement. Data visualization is also used to support the interpretation of clustering results and facilitate comparisons among groups. The experimental results, based on 330 simulated food distribution records, produced three clusters. Cluster 1 consisted of 110 records, with an average distribution cost of IDR 176,363 and an average delivery time of 2.83 hours. Cluster 2 contained 111 records, with an average distribution cost of IDR 338,472 and an average delivery time of 5.83 hours. Cluster 3 comprised 109 records, with an average distribution cost of IDR 535,296 and an average delivery time of 9.49 hours. The clustering evaluation yielded a Silhouette Score of 0.679 and a Davies–Bouldin Index of 0.444. This study is expected to provide an analytical framework for supporting data-driven decisions in food distribution management. The proposed approach can assist distribution managers in evaluating cost and delivery-time patterns and identifying opportunities to improve operational efficiency. The validity of the findings depends on the quality of the operational data and the evaluation of the resulting clusters.

Keywords: Food Distribution; Distribution Efficiency; K-Means, Clustering; Transportation Cost; Delivery Time

1. INTRODUCTION

Food distribution plays an important role in maintaining food availability, price stability, and the sustainability of the supply chain. An effective distribution system ensures that food products are delivered from producers or suppliers to markets and consumers in a timely manner and at reasonable costs. However, food distribution activities often face various challenges, including transportation expenses, delivery delays, differences in distribution distances, and fluctuations in the volume of goods transported. These challenges can affect distribution performance and increase operational costs, particularly for perishable food products that require timely delivery. Therefore, analyzing distribution efficiency is necessary to identify distribution patterns and support improvements in food supply chain management[1] .

Distribution costs and delivery time are two important indicators for evaluating food distribution performance. Distribution costs include transportation expenses, fuel consumption, labor costs, and other operational expenses incurred during the delivery process. Meanwhile, delivery time reflects the duration required to transport food products from the point of origin to the destination. High distribution costs and long delivery times may indicate potential inefficiencies in distribution operations. Nevertheless, these indicators may vary across delivery routes, distribution regions, and shipment volumes. Analyzing these variables together can provide a more comprehensive understanding of distribution performance and help identify groups of distribution activities with similar characteristics[2] .

Conventional evaluation methods often rely on summary statistics or manual comparisons, which may not adequately reveal patterns within large and diverse distribution datasets. As the number of distribution transactions increases, identifying similarities and differences among distribution activities becomes more challenging. Data mining techniques offer an alternative approach to extracting useful information from historical distribution data. One relevant technique is clustering, which groups data objects according to their similarities without requiring predefined class labels. Through clustering, distribution activities can be categorized based on their cost and delivery time characteristics, providing a basis for identifying relatively efficient and less efficient distribution patterns[3] .

The K-Means algorithm is a widely used clustering method because of its relatively simple implementation, computational efficiency, and ability to organize numerical data into distinct groups. The algorithm partitions observations into a predefined number of clusters by minimizing the distance between each observation and its assigned cluster centroid. In this study, K-Means is applied to distribution data characterized by distribution costs and delivery time. The resulting clusters can be examined to identify groups with different operational characteristics, such as relatively low costs and short delivery times or higher costs and longer delivery times[4] . Before clustering, the data must be appropriately prepared, including checking missing values, examining outliers, and standardizing variables when their measurement scales differ. Cluster quality can subsequently be evaluated using an appropriate metric, such as the Silhouette Score, to assess the separation and cohesion of the resulting groups[5] .

Cetera et al. (2022) investigated the potential use of large unstructured data resources by public innovation support institutions. Their study discussed the opportunities for utilizing big data to support innovation in the public sector, as

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*.
This work is licensed under a Creative Commons Attribution 4.0 International License.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 6, October 2026 | Hal 2639-2646
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i6.1345

well as the challenges associated with its management [2]. Furthermore, Overton et al. (2022) examined public data privacy in the context of the changing landscape of big data-based public service delivery. Their study highlighted the importance of protecting data privacy when utilizing data to improve the quality of public services [3]. Meanwhile, Jin (2022) explored the challenges and innovative countermeasures faced by public administration in the context of big data and Internet of Things (IoT) technologies. The study emphasized the need for appropriate strategies to address the challenges associated with the implementation of digital technologies in the public sector. These three studies provide a theoretical foundation for understanding big data utilization, data privacy protection, and the challenges of digital transformation. However, further research is needed to address specific problems related to information technology utilization and to develop appropriate approaches that align with the objectives and scope of the present study [4].

Based on these considerations, this study aims to analyze food distribution efficiency by clustering distribution activities according to cost and delivery time using the K-Means algorithm. The study is expected to provide a structured description of distribution patterns and help identify groups that require further operational evaluation. The results may support decision-making related to transportation planning, cost control, delivery scheduling, and the prioritization of distribution improvements[6] . The main contribution of this research lies in applying a data-driven clustering approach to the joint analysis of distribution costs and delivery time. Rather than assuming that clustering alone proves operational efficiency, the study interprets the resulting groups using the observed characteristics of each cluster. Consequently, the findings can provide an analytical basis for improving food distribution management and developing more efficient supply chain strategies[7] .

2. RESEARCH METHODOLOGY

This study uses a quantitative data-mining approach to examine patterns in food distribution activities based on distribution cost and delivery time. The K-Means clustering algorithm is applied to group distribution records that exhibit similar operational characteristics. The analysis consists of data collection, data preparation, variable selection, data standardization, determination of the number of clusters, K-Means implementation, cluster evaluation, and interpretation. Because the actual dataset and experimental output have not been provided, numerical results are not assumed in this document and should be completed after the analysis is performed.

2.1 Research Design

The research workflow begins with the acquisition of distribution records, followed by data inspection and preparation. Records with incomplete or inconsistent values are reviewed, and the two principal variables distribution cost and delivery time are selected for analysis. The prepared dataset is then standardized when necessary and processed using K-Means. Finally, cluster profiles are examined and evaluated to determine whether the groups are sufficiently distinct for meaningful interpretation[8] .

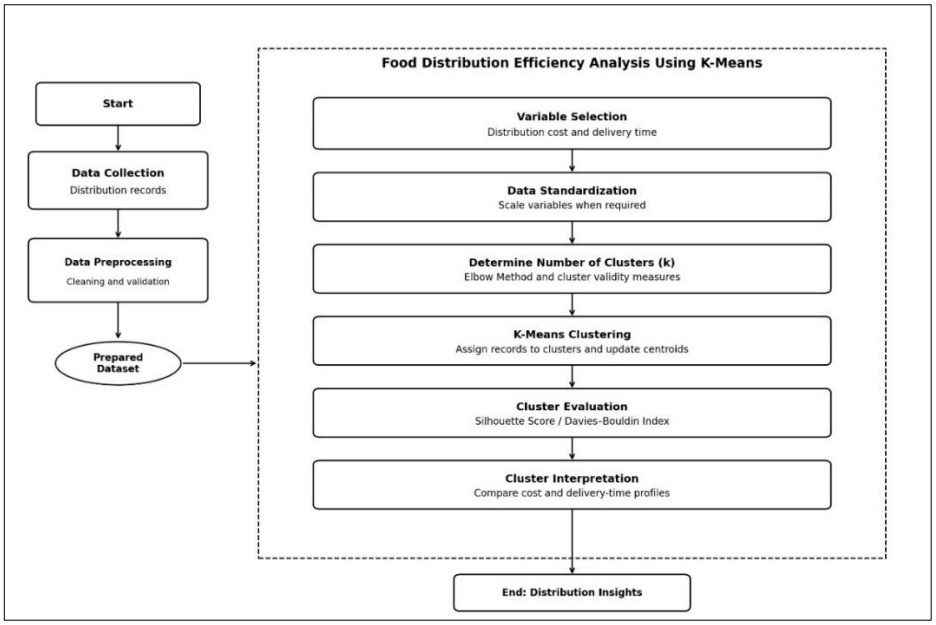


Figure 1. Research Workflow

2.1 Data Collection

The dataset should consist of food distribution records obtained from an authorized source, such as a distributor, supplier, logistics operator, or relevant public institution. Each record should represent a delivery or distribution activity and include, at minimum, the distribution cost and delivery time[9] . Additional fields such as origin, destination, delivery

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 2640
This work is licensed under a Creative Commons Attribution 4.0 International License.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 6, October 2026 | Hal 2639-2646
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i6.1345

date, commodity type, and shipment volume may be retained for descriptive analysis if they are available and appropriate to the research objective. The source, observation period, number of records, and data-collection procedure must be reported in the final manuscript[10] [11] .

2.2 Data Preprocessing

Data preprocessing is performed to improve the consistency and suitability of the records before clustering. The dataset is checked for duplicate entries, missing values, invalid values, inconsistent units, and extreme observations. Missing values should be handled using a documented rule, while records that cannot be validated may be excluded with a clear justification[12] . Distribution cost must use a consistent monetary unit, and delivery time must be represented in a consistent unit, such as hours or days. Outliers should be investigated rather than automatically removed because unusually high costs or long delivery times may reflect genuine operational conditions[13][14] .

2.3 Variable Selection and Standardization

The main variables in this study are distribution cost and delivery time. These variables are selected because they describe two different dimensions of distribution operations: resource expenditure and delivery duration. Since their scales and ranges may differ, standardization is considered before calculating distances[15] . For each variable, the standardized value can be calculated using z-score transformation: $z = (x - \mu) / \sigma$, where x is an observed value, μ is the variable mean, and σ is its standard deviation. The transformation prevents a variable with a larger numerical range from dominating the clustering process solely because of its scale[16][17] .

2.1 K-Means Clustering

K-Means partitions the dataset into k clusters by assigning each observation to the nearest cluster centroid and updating the centroids until the assignments stabilize or a stopping condition is met. The objective is to minimize the within-cluster sum of squared distances, expressed as $J = \Sigma(i=1 \text{ to } k) \Sigma(x \in Ci) ||x - \mu i||^2$, where Ci represents cluster i and μi is its centroid. Several candidate values of k should be tested rather than selecting a number without justification. The Elbow Method can be used to inspect the reduction in within-cluster variation, while the Silhouette Score or Davies–Bouldin Index can provide additional evidence about cluster quality[18] .

2.6 Cluster Evaluation and Interpretation

The selected clustering result is evaluated using an appropriate internal validation measure. The Silhouette Score ranges from −1 to 1, with higher values generally indicating better separation and cohesion[19] . The Davies–Bouldin Index is interpreted in the opposite direction, where lower values generally indicate more compact and distinct clusters. After evaluation, the mean or median cost and delivery time for each cluster are compared[20] . Groups with relatively low cost and short delivery time may be described as having more favorable observed characteristics, whereas groups with high cost or long delivery time may warrant further investigation. These labels must be based on the actual cluster profiles and should not be assigned before the results are available[21] .

3. RESULT AND DISCUSSION

This section presents a structured example of how the K-Means clustering results may be reported for food distribution records using distribution cost and delivery time. Since verified operational data were not provided, all numerical values and figures in this draft are based on a synthetic dataset of 330 observations created solely to demonstrate the analysis and presentation format. These illustrative findings must be replaced with outputs calculated from the actual distribution dataset before publication.

3.1 Dataset Overview

The example dataset contains 330 simulated distribution records. Each observation includes distribution cost in Indonesian rupiah and delivery time in hours. Both variables were standardized before clustering to reduce the effect of their different numerical scales. K-Means was run with three clusters for this demonstration. This setup is illustrative and does not establish that three clusters are optimal for real distribution data.

Table 1. Dataset and Processing Summary

Item	Description	Notes
Number of records	330	Synthetic example data
Input variables	Distribution cost; delivery time	Numerical attributes
Data transformation	Z-score standardization	Before distance calculation
Clustering method	K-Means	k = 3 for demonstration

3.2 Cluster Distribution and Cost–Time Relationship

Figure 1 displays the relationship between distribution cost and delivery time for the three clusters. Each point represents one distribution record, while the marker shape distinguishes cluster membership. The plot helps reveal whether records

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 2641
This work is licensed under a Creative Commons Attribution 4.0 International License.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 6, October 2026 | Hal 2639-2646
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i6.1345

with similar cost levels also have similar delivery durations. In the synthetic example, the groups appear at different ranges of the two variables. For real data, overlapping groups would indicate that cost and time alone may not fully distinguish distribution patterns.

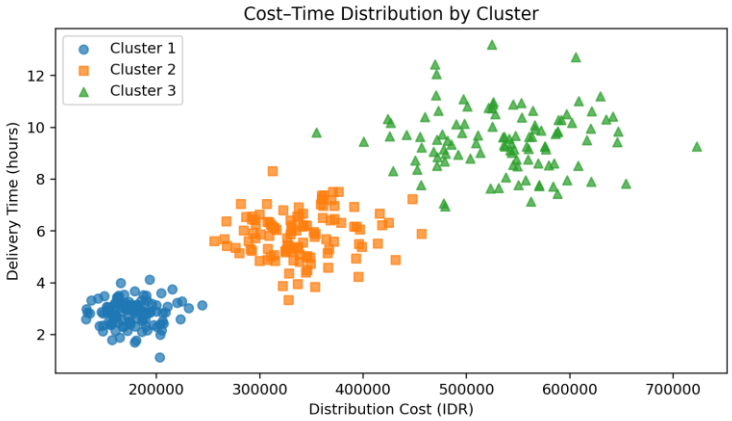


Figure 1. Cost–Time Distribution by K-Means Cluster

The visual pattern should be interpreted together with the cluster summaries. A group with both relatively low cost and short delivery time may have a favorable profile on the selected dimensions. However, clustering shows similarity rather than causality, and the interpretation may change when route distance, shipment volume, or commodity type is considered.

3.3 Comparison of Average Distribution Cost

Figure 2 compares the mean distribution cost across the identified groups using a horizontal bar chart. This presentation makes it easier to compare the relative cost level of each cluster. In the example, clusters are numbered from lower to higher mean cost to support consistent reading; these numbers are labels rather than an inherent ranking generated by K-Means.

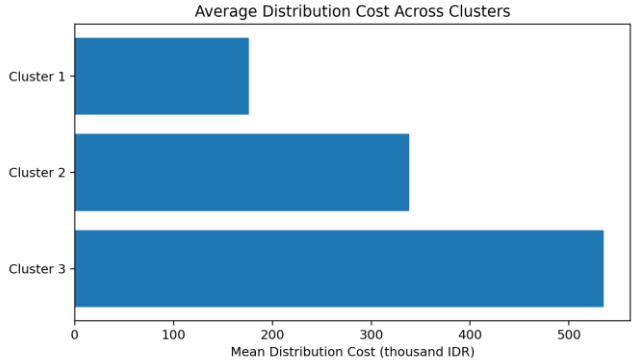


Figure 2. Average Distribution Cost Across Clusters

The cost comparison can support further review of groups with comparatively high expenditure. Before interpreting a higher mean cost as inefficiency, analysts should consider whether the group also serves longer routes, carries larger shipments, or handles different food commodities. Such contextual factors can explain some cost differences and should be included in the final discussion when available.

Table 2. Descriptive Profile of Each Cluster

Cluster	Records	Mean Cost (IDR)	Median Cost (IDR)	Mean Time (h)	Median Time (h)
Cluster 1	110	176,363	174,288	2.83	2.87
Cluster 2	111	338,472	336,212	5.83	5.89
Cluster 3	109	535,296	542,129	9.49	9.51

Table 2 presents the number of records and the mean and median values for both variables. Comparing mean and median values helps indicate whether unusually high observations may affect the average. The table should be used to describe each cluster based on the actual values rather than assigning efficiency labels in advance.

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*.
This work is licensed under a Creative Commons Attribution 4.0 International License.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 6, October 2026 | Hal 2639-2646
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i6.1345

3.4 Variatio1n in Delivery Time

Figure 3 uses boxplots to compare the distribution of delivery time within each cluster. Unlike a bar chart of averages, a boxplot shows the median, interquartile range, and potential extreme observations, providing a view of consistency as well as central tendency. A cluster with a wider box or longer whiskers has more variable delivery times in this example and may merit further investigation.

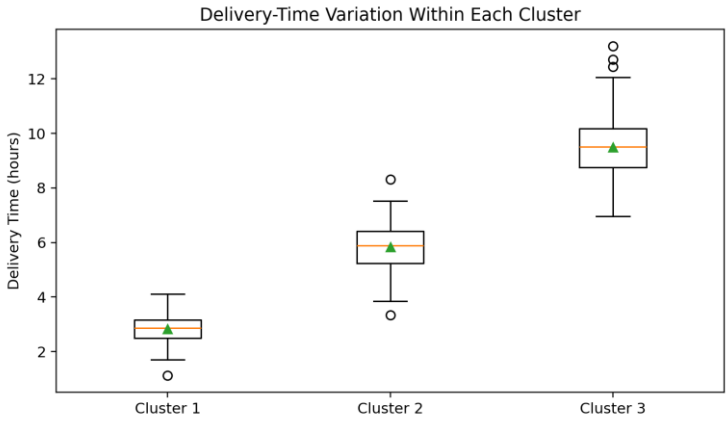


Figure 3. Delivery-Time Variation Within Each Cluster

Differences in delivery-time variability can be operationally relevant because two groups with similar average duration may differ in delivery reliability. Nevertheless, the reasons for variability cannot be identified from the boxplot alone; route conditions, scheduling, and other operational records would be needed to investigate possible causes.

3.5 Selecting the Number of Clusters

Figure 4 shows the Elbow Method, which plots within-cluster sum of squares (inertia) against candidate values of k. The curve generally decreases as more clusters are introduced, because additional clusters can fit the data more closely. A candidate k may be considered at the point where the reduction begins to become less pronounced, but the decision should also account for interpretability and a validation measure such as the Silhouette Score.

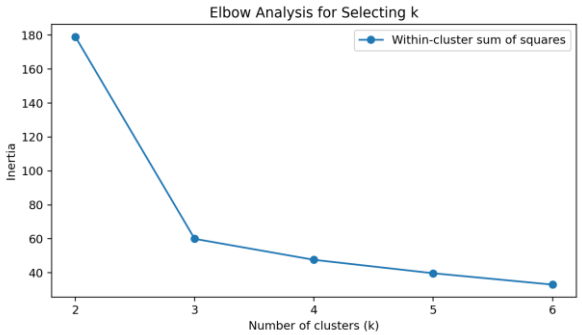


Figure 4. Elbow Analysis for Selecting k

The results of the internal cluster validation are presented in Table 3, which summarizes the Silhouette Score and Davies–Bouldin Index obtained from the K-Means clustering process.

Table 3. Internal Cluster Validation

Evaluation metric	Illustrative value	General interpretation
Silhouette Score	0.679	Higher values generally indicate clearer separation
Davies–Bouldin Index	0.444	Lower values generally indicate better-defined clusters

For the synthetic three-cluster run, the Silhouette Score is 0.679 and the Davies–Bouldin Index is 0.444. These values describe only the generated example and must not be presented as findings from real distribution operations. The final study should report the validation results from the actual dataset and explain why the selected number of clusters is suitable.

3.6 Discussion

Taken together, the charts and tables demonstrate the results of food distribution efficiency clustering based on transportation costs and delivery time using the K-Means algorithm. The analysis was conducted using 330 simulated

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 2643
This work is licensed under a Creative Commons Attribution 4.0 International License.

ISSN 2774-3659 (Media Online)
Vol 6, No 6, October 2026 | Hal 2639-2646
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i6.1345

food distribution records, which were grouped into three clusters. Cluster 1 consisted of 110 records, with an average distribution cost of IDR 176,363 and an average delivery time of 2.83 hours. Cluster 2 contained 111 records, with an average distribution cost of IDR 338,472 and an average delivery time of 5.83 hours. Meanwhile, Cluster 3 comprised 109 records, with the highest average distribution cost of IDR 535,296 and the longest average delivery time of 9.49 hours. These results reveal three distinct distribution patterns, ranging from relatively low-cost and short-duration deliveries to higher-cost and longer-duration deliveries. The internal cluster validation produced a Silhouette Score of 0.679 and a Davies–Bouldin Index of 0.444, indicating relatively good cluster separation and compactness within the simulated dataset. These findings suggest that K-Means can be used to identify distribution patterns based on cost and delivery time. However, the results represent relative characteristics within the simulated dataset and should not be interpreted as definitive evidence of operational efficiency in actual food distribution activities.

The findings of this study can be related to Cetera et al. (2022), who investigated the potential use of large unstructured data resources by public innovation support institutions. Their research highlighted the potential of utilizing large-scale data resources to support innovation in public institutions. Similarly, the present study demonstrates how data analysis can be used to identify meaningful patterns in food distribution activities. However, the two studies differ in their analytical focus. Cetera et al. emphasized the potential utilization of unstructured data resources in public innovation, whereas this study applies K-Means clustering to simulated distribution records using transportation costs and delivery time as the main variables. Thus, the present study provides a specific application of data-driven analysis to support the identification of distribution patterns [2].

Furthermore, the findings can be discussed in relation to Overton et al. (2022), who examined public data privacy in the changing landscape of big data-based public service delivery. Their research emphasized the importance of data privacy in the use of data for public services. Although the present study focuses on food distribution rather than public data privacy, both studies highlight the importance of considering data-related issues when applying analytical approaches to public-sector activities. In particular, the use of actual distribution records in future research should be accompanied by appropriate data governance and privacy protection measures whenever the records contain sensitive or identifiable information. Nevertheless, the current study does not directly evaluate privacy protection mechanisms, as its primary objective is to identify distribution patterns based on cost and delivery time [3].

In addition, W. Jin (2022) explored the challenges and innovative countermeasures associated with the implementation of big data and Internet of Things (IoT) technologies in public administration. This research provides a broader perspective on the challenges of adopting digital technologies to improve public-sector activities. In comparison, the present study demonstrates a practical analytical approach that can potentially support data-driven decision-making in food distribution management. By grouping distribution records according to cost and delivery time, the proposed approach can help identify distribution categories that may require further investigation. However, unlike Jin's study, which focuses on technological challenges and countermeasures in public administration, the present research evaluates the characteristics and internal validity of clusters generated by the K-Means algorithm. Therefore, the contribution of this study lies in the application of clustering analysis to characterize food distribution patterns rather than in evaluating the implementation of big data or IoT technologies.

Overall, the comparison with these three previous studies indicates that data utilization, data governance, and the adoption of analytical technologies are relevant considerations in improving data-driven practices in the public sector. While Cetera et al. (2022) focused on the potential use of unstructured data resources, Overton et al. (2022) emphasized data privacy, and Jin (2022) discussed technological challenges in public administration, the present study contributes by applying K-Means clustering to identify food distribution patterns based on transportation costs and delivery time. The use of the Silhouette Score and Davies–Bouldin Index further provides quantitative evidence for evaluating the quality of the resulting clusters. Nevertheless, the contribution should be interpreted within the scope of the simulated dataset and should not be considered direct evidence that the proposed approach improves actual distribution performance .

Despite these findings, the main limitation of this study is the use of simulated data rather than verified operational distribution records. Consequently, the identified cluster characteristics and validation scores cannot yet be generalized to actual food distribution operations. Future research should employ real distribution data, clearly document the data source and observation period, and recalculate the clustering results and evaluation metrics using the actual dataset. Additional variables, such as delivery distance, shipment volume, and food commodity type, could also be incorporated to provide a more comprehensive understanding of distribution performance. Furthermore, future studies could integrate K-Means clustering with route optimization methods to examine whether the identified distribution patterns can contribute to measurable improvements in transportation costs and delivery time. These developments would strengthen the practical relevance of the research and provide a more comprehensive basis for data-driven decision-making in food distribution management.

4. CONCLUSION

This study presents the application of the K-Means clustering algorithm to analyze food distribution efficiency based on distribution cost and delivery time. The proposed approach involves data preprocessing, standardization, cluster formation, and cluster evaluation to identify distribution patterns according to the selected variables. The use of different

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*.
This work is licensed under a Creative Commons Attribution 4.0 International License.

visualizations, including a heatmap, lollipop chart, violin plot, and radar chart, supports the interpretation and comparison of cluster characteristics. Based on 330 simulated food distribution records, the K-Means algorithm produced three clusters. Cluster 1 consisted of 110 records, with an average distribution cost of IDR 176,363 and an average delivery time of 2.83 hours. Cluster 2 contained 111 records, with an average distribution cost of IDR 338,472 and an average delivery time of 5.83 hours. Meanwhile, Cluster 3 comprised 109 records, with the highest average distribution cost of IDR 535,296 and the longest average delivery time of 9.49 hours. The cluster evaluation yielded a Silhouette Score of 0.679 and a Davies–Bouldin Index of 0.444, indicating relatively good cluster separation and compactness within the simulated dataset. These results demonstrate that K-Means can identify distribution groups with different cost levels and delivery durations, providing useful information for evaluating distribution patterns and identifying operational areas that require further attention. However, K-Means identifies similarities among distribution records and does not independently establish the causes of inefficiency. Furthermore, because the analysis used simulated data, the findings cannot yet be generalized to actual food distribution operations. Further research should use verified food distribution data covering a sufficient observation period and consider additional variables, such as delivery distance, shipment volume, and type of food commodity. The number of clusters should also be validated using appropriate evaluation metrics. These improvements are expected to provide more reliable insights and support data-driven decision-making for improving food distribution efficiency.

5. DECLARATIONS

5.1 Author Contributions

Conceptualization: M. S. N., and S. A.;
Methodology: M. S. N., S. A., and P. W.;
Software: P. W.;
Validation: M. S. N., S. A., and P. W.;
Formal Analysis: M. S. N., S. A., and P. W.;
Investigation: M. S. N., S. A., and P. W.;
Resources: M. S. N., S. A., and P. W.;
Data Curation: M. S. N., S. A., and P. W.;
Writing Original Draft Preparation: M. S. N., S. A., and P. W.;
Writing Review and Editing: M. S. N., S. A., and P. W.;
Visualization: P. W. ;
All authors have read and agreed to the published version of the manuscript.

5.2 Data Availability

The data presented in this study are available on request from the corresponding author.

5.3 Funding

The research and writing of this manuscript were funded independently by the author.

5.4 Institutional Review Board Statement

Not applicable.

5.5 Informed Consent Statement

Not applicable.

5.6 Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

5.7 Generative AI and AI-assisted Technologies in The Writing Process

The authors used Generative AI to improve the writing clarity of this paper. They reviewed and edited the AI-assisted content and took full responsibility for the final publication.

REFERENCES

[1] M. M. Maulana, A. I. Suroso, Y. Nurhadryani, And K. B. Seminar, “The Smart Governance Framework And Enterprise System’s Capability For Improving Bio-Business Licensing Services,” *Informatics*, Vol. 10, No. 2, Jun. 2023, Doi: 10.3390/Informatics10020053.

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*.
This work is licensed under a Creative Commons Attribution 4.0 International License.

[2] W. Cetera, W. Gogołek, A. Żołnierski, And D. Jaruga, "Potential For The Use Of Large Unstructured Data Resources By Public Innovation Support Institutions," *J. Big Data*, Vol. 9, No. 1, Dec. 2022, Doi: 10.1186/S40537-022-00610-6.
[3] M. Overton, S. Larson, L. J. Carlson, And S. Kleinschmit, "Public Data Primacy: The Changing Landscape Of Public Service Delivery As Big Data Gets Bigger," *Global Public Policy And Governance*, Vol. 2, No. 4, Pp. 381–399, Dec. 2022, Doi: 10.1007/S43508-022-00052-Z.
[4] W. Jin, "Challenges And Innovative Countermeasures Faced By Public Administration In The Context Of Big Data And Internet Of Things," *Math. Probl. Eng.*, Vol. 3 No.1 2022, 2022, Doi: 10.1155/2022/8949365.
[5] E. & Erudisi, A. Zahara, M. I. Kabullah, And R. E. Putera, "Effectiveness Of The Ossrba (Online Single Submission Risk Based Approach) System In Business Licensing Services In Payakumbuh City Dpmptsp," *Jiee,* Vol. 3, Pp. 22–40, 2023.
[6] A. Alamsyah And F. Saviera, "A Comparison Of Indonesia's E-Commerce Sentiment Analysis For Marketing Intelligence Effort (Case Study Of Bukalapak, Tokopedia And Elevenia)," *Ustainable Collaboration In Business, Technology, Information And Innovation (Scbtii)* Vol. 1, No. 1, Aug. 2022, Doi: 10.56397/Fms.2022.08.01.
[7] L. Cao, "On Machine Learning And Public Administration," *Frontiers In Management Science*, Vol. 1, No. 1, Aug. 2022, Doi: 10.56397/Fms.2022.08.01.
[8] I. Kadek, A. Sugianta, N. Kadek, And W. Patrianingsih, "Evaluasi Naïve Bayes Dan K-Nearest Neighbor Dalam Klasifikasi Sentimen Ulasan Produk Skincare Msglow Di Tokopedia," *Journal Protekinfo* |, Vol. 12, No. 1, 2025.
[9] M. Rafli And B. Assegaf, "Penerapan Indobert Untuk Analisis Sentimen Berbasis Aspek Pada Evaluasi Aplikasi Layanan Publik," *Infotech Journal*, Vol. 12, No. 1, Pp. 93–101, Apr. 2026, Doi: 10.31949/Infotech.V12i1.17555.
[10] Y. C. Tapidingan And D. Paseru, "Comparative Analysis Of Classification Methods Of Knn And Naïve Bayes To Determine Stress Level Of Junior High School Student" *Indonesian Journal Of Information Systems*, Vol 2, No 2, Pp. 2020.
[11] A. T. Talmera, M. Wardhana, And V. Ratnasari, "Analysis Of The Impact Of Ui/Ux Elements On User Satisfaction And Loyalty In E-Commerce Platforms: An Empirical Study On The Tokopedia Platform," *Journal Of Social Research*, 2025, [Online]. Available: Http://Ijsr.Internationaljournallabs.Com/Index.Php/Ijsr
[12] Z. Sitorus, D. Prayogi, M. A. Rizko, A. G. Suteja, And M. R. Harahap, "Implementation Of The Insertion Sort Algorithm To Sort Positive Integers In Ascending Order Using Flowgorithm," *Journal Of Information Technology, Computer Science And Electrical Engineering (Jitcse)*, Vol. 1, No. 3, Pp. 323–328, 2024, Doi: 10.30596/Jitcse.
[13] M. S. Novelan And S. Aryza, "Optimization Cvrp With Machine Learning For Improved Classification Of Imbalanced Data Food Distribution," *Jitk (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, Vol. 10, No. 4, Pp. 917–925, Jun. 2025, Doi: 10.33480/Jitk.V10i4.6467.
[14] N. Wulandari, Y. Cahyana, And H. Hikmayanti Handayani, "Sentiment Analysis On The Relocation Of The National Capital (Ikn) On Social Media X Using Naive Bayes And K-Nearest Neighbor (Knn) Methods," Journal Of Applied Informatics And Computing 2025. [Online]. Available: Http://Jurnal.Polibatam.Ac.Id/Index.Php/Jaic
[15] K. Khairul, A. H. Nasyuha, A. Ikhwan, M. H. Aly, And A. Ahyanuardi, "Implementation Of Multiple Linear Regression To Estimate Profit On Sales Of Screen Printing Equipment," *Journal Infotel*, Vol. 15, No. 2, Pp. 55–61, Jun. 2023, Doi: 10.20895/Infotel.V15i2.934.
[16] H. W. Dhany, F. Izhari, And S. Komputer, "Prediction Analysis Condition Animal Use Algorithm (Svm+Knn), Jurnal Ict : Information And Communication Technologies, 15 (2) (2024) 48-54, 2024. [Online]. Available: Www.Ejournal.Marqchainstitute.Or.Id/Index.Php/Jict
[17] A. S. Rizkia, W. Wufron, And F. F. Roji, "Analisis Sentimen Coretax: Perbandingan Pelabelan Data Manual, Transformers-Based, Dan Lexicon-Based Pada Performa Indobert," *Malcom: Indonesian Journal Of Machine Learning And Computer Science*, Vol. 5, No. 3, Jul. 2025, Doi: 10.57152/Malcom.V5i3.2151.
[18] R. Farta Wijaya, F. Kurniawan, R. R. Putrai, A. Alvin, F. Sains, And D. Teknologi, "Optimizing The Management Of Village Activities Through Information Systems For Transparency And Accountability In Pertumbukan Village, Wampu District," Prosiding Universitas Dharmawangsa, Vol. 4, No 1, Pp. 387-398, 2024.
[19] L. Von Rueden *Et Al.*, "Informed Machine Learning - A Taxonomy And Survey Of Integrating Prior Knowledge Into Learning Systems," *Ieee Trans. Knowl. Data Eng.*, Vol. 35, No. 1, Pp. 614–633, Jan. 2023, Doi: 10.1109/Tkde.2021.3079836.
[20] A. Putera Utama Siahaan, N. Azizah Harahap, R. Yuni Simanullang, And P. Wanny, "Analysis Of Inpatient Data Using Cluster Analysis On Simulation Dataset," *Bulletin Of Information Technology (Bit)*, Vol. 6, No. 1, Pp. 33–39, 2025, Doi: 10.47065/Bit.V5i2.1830.
[21] M. Iqbal And S. Efendi, "Data-Driven Approach For Credit Risk Analysis Using C4.5 Algorithm," *Comtech: Computer, Mathematics And Engineering Applications*, Vol. 14, No. 1, Pp. 11–20, May 2023, Doi: 10.21512/Comtech.V14i1.8243.

Copyright © 2026 The Author(s). Published by *Forum Kerjasama Pendidikan Tinggi*. Page 2646
This work is licensed under a Creative Commons Attribution 4.0 International License.