Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 5, **Number** 2, November 2025
https://doi.org/10.47709/brilliance.v5i2.6992

Analysis Of Tokopedia Product Clustering Using The K-Means And K-Medoids Algorithms

Raihan Malik[1*], Pradita Eko Prasetyo Utomo[2], Benedika Ferdian Hutabarat[3]
[1,2,3]Universitas Jambi, Indonesia
[1] raihanmalik00@gmail.com, [2] pradita.eko@unja.ac.id, [3]benedika@unja.ac.id

open access

***Corresponding Author**

Article History:
Submitted: 23-09-2025
Accepted: 01-10-2025
PUBLISHED: 10-10-2025

Keywords:
Clustering; K-Means; K-Medoids; Silhouette Score; Davies-Bouldin Index.

Brilliance: Research of Artificial Intelligence is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

ABSTRACT

The Indonesian e-commerce market has experienced extraordinary growth, driven by increasing internet penetration and smartphone adoption, which necessitates advanced data analysis for competitive advantage. Clustering is a crucial data mining technique used to group products based on similar characteristics, providing in-depth insights into product performance. Previous studies often focused on single performance metrics, overlooking the nuances of combining multiple variables. This study aims to address this gap by implementing and comparing the K-Means and K-Medoids clustering algorithms on Tokopedia product data using a combination of numerical attributes: Price, Customer Rating, Number Sold, and Total Review. The methodology involved data preprocessing, Min-Max Scaling for normalization, and using the Elbow Method to determine the optimal number of clusters, which was found to be K=2. The clustering quality was rigorously evaluated using the Davies-Bouldin Index (DBI) and Silhouette Score. The results demonstrate that K-Means exhibits superior performance, achieving a lower DBI of 0.5717 and a higher Silhouette Score of 0.6012, compared to K-Medoids (DBI: 0.5870; Silhouette Score: 0.5857). Furthermore, K-Means proved significantly more efficient computationally, with an execution time of 0.0947 seconds versus 0.1622 seconds for K-Medoids. The main conclusion is that K-Means is more effective in creating compact and clearly separated clusters. This research contributes a valuable analytical framework for e-commerce managers to comprehensively understand product profiles, guiding more effective marketing and recommendation strategies.

INTRODUCTION

The e-commerce market in Indonesia has experienced tremendous growth over the past decade, driven by increasing internet penetration and smartphone adoption (V. K. Sari & Nasution, 2024). Increasingly inclusive digital access allows consumers from even remote areas to shop online with ease. Major e-commerce platforms such as Tokopedia and Shopee have played a significant role in the digitalization of Micro, Small, and Medium Enterprises (MSMEs), contributing significantly to their revenue and market reach (Purba et al., 2025). Overall, this trend has not only changed people's consumption patterns but also propelled Indonesia's digital economy to become a major force in Southeast Asia.

In a competitive e-commerce ecosystem, platform providers compete not only on the number of products or prices, but also on the quality of user experience and personalized services. The use of AI-based automated recommendation systems has become a crucial strategy for increasing user interaction and driving sales. One of the most commonly used methods to achieve this personalization is clustering, a data analysis technique that groups data based on similar characteristics without requiring specific class labels (Gymnastiar & Bahtiar, 2024).

Although clustering has been widely applied in e-commerce, most previous studies tend to focus on only one performance attribute, such as sales levels (Harjono et al., 2023) or product ratings (Rahman & Suroyo, 2021). This gap ignores the potential insights that can be gained from multivariate analysis, where high ratings do not always correlate with high sales (Rahman & Suroyo, 2021). For example, a product may have a perfect rating but very low sales, or vice versa. Therefore, the lack of a holistic approach that integrates several product performance attributes, such as ratings and sales volume simultaneously, is a critical gap in e-commerce literature. This study aims to address this gap by exploring how the combination of these attributes can produce more relevant and nuanced product groupings.

As part of the solution, this study implements and compares two popular clustering algorithms, namely K-Means and K-Medoids. By combining both variables, this study is expected to identify a more holistic product grouping. The main contribution of this study is to provide a new framework for e-commerce managers to understand product performance more comprehensively, which can ultimately guide more effective marketing strategies and recommendation systems.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

E-ISSN : 2807-9035
Volume 5, **Number** 2, November 2025
https://doi.org/10.47709/brilliance.v5i2.6992

LITERATURE REVIEW

Clustering is a data mining technique used to group data into clusters based on similar characteristics. This method falls under the category of unsupervised learning because the clustering process does not require class labels on the analyzed data (Hermawati et al., 2020). In general, clustering is divided into two categories: hierarchical and non-hierarchical. Hierarchical clustering forms a hierarchical structure, while non-hierarchical methods directly group data into a number of predefined clusters (Saputri & Arianto, 2023). The use of non-hierarchical methods is more efficient for large datasets, while hierarchical methods tend to be sensitive to noise and outliers.

In the clustering approach, two popular algorithms used are K-Means and K-Medoids (Hermawati et al., 2020). The K-Means method groups data by considering the proximity between data based on Euclidean distance, while the K-Medoids algorithm focuses on finding medoids in each cluster, which are points that represent the center of the cluster (Arora et al., 2016). K-Means is known to be accurate and simple in its application, but it is limited because it requires a predetermined number of clusters and is less capable of detecting outliers. Conversely, K-Medoids is more robust to outliers, but does not fully represent the influence of all elements in the cluster and still depends on the predetermined number of clusters (Riyahi & Martín, 2025). The selection of the appropriate method can be determined based on data characteristics and evaluation using validity metrics. To obtain quality cluster results, the Elbow Method approach will be used by comparing the proximity between elements in one cluster to their distance from elements in different clusters (Riyahi & Martín, 2025). The data is first standardized, then the optimal number of clusters is determined using the Elbow method, which considers data variability (Lukáč et al., 2025).

Validation of the results produced by the clustering algorithm is an important aspect of the overall clustering process (Arbelaitz, 2013). Therefore, in this study, the metrics commonly used in the evaluation are the Davies-Bouldin Index (DBI) and Silhouette Score. These two metrics have been proven effective in measuring the closeness and separation between data groups (Syahkur & Hartama, 2024). DBI measures the extent to which clustering results are able to distinguish between different data groups while maintaining the closeness of the data to its cluster center (Umagapi & Umaternate, 2023). The lower the DBI value, the better the cluster formation quality (Idrus et al., 2022). Silhouette Score assesses the accuracy of data points in clusters, with values close to 1 indicating good clustering quality (Rhomadhona et al., 2025). The higher the Silhouette Coefficient value, the clearer the cluster definition in the model (Kulkanjanapiban & Silwattananusarn, 2025).

METHOD

The flow of this research will be visualized with the following fishbone diagram:

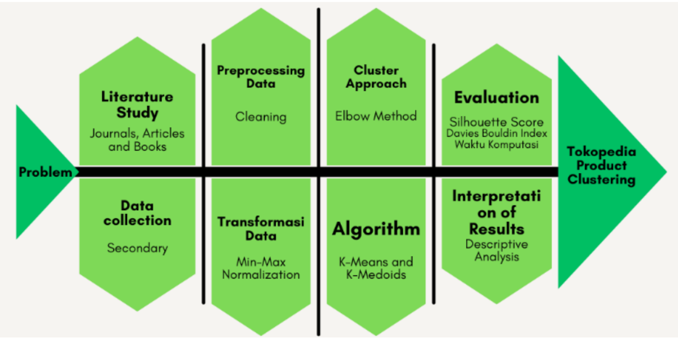

Fig 1. Fishbone Diagram Alur Penelitian

Data Collection

Data collection in this study was conducted using secondary data sources from the PRDECT-ID dataset (Product Reviews Dataset for Emotion Classification Tasks – Indonesian). This dataset was developed by Sutoyo et al, 2022 and published through the Mendeley Data platform in 2022. The dataset contains 5,401 product review entries taken directly from the Tokopedia e-commerce platform. The available data covers various important attributes such as Category, Product Name, Location, Price, Overall Rating, Number Sold, Total Review, Customer Rating, Customer Review, Sentiment, and Emotion. In this study, the analysis focuses on numerical attributes such as Price, Customer Rating, Number Sold, and Total Review, as these attributes are considered most relevant in forming product segmentation based on sales performance and consumer perception. The category attribute (Category) is still used in the interpretation of the clustering results to identify product distribution based on business domain.

Data Preprocessing

Data preprocessing in this study was carried out systematically to ensure optimal data quality prior to clustering. The process began with an initial data assessment to identify problems such as missing values, duplicates, text inconsistencies, and extreme data. Next, text standardization was performed on product names and categories, followed

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 5, **Number** 2, November 2025
https://doi.org/10.47709/brilliance.v5i2.6992

by mapping specific categories into eight main groups to simplify analysis. Incomplete or duplicate data was then cleaned, and categories with insufficient data were removed to maintain balance. Outliers were handled in numerical attributes (Price, Number Sold, and Total Review) using the Interquartile Range (IQR) method, while customer ratings were excluded because of their fixed scale. Finally, relevant attributes were selected and prepared for scale transformation using Min-Max Scaling, ensuring the data was ready for accurate clustering analysis. All stages were implemented using Python on Google Colab, with the help of the pandas, numpy, and openpyxl libraries.

Data Transformation

This transformation aims to equalize the scale between numerical attributes so that no single attribute dominates the distance calculation in the clustering process. For this purpose, the Min-Max Scaling technique is used, which is a normalization method that converts the original attribute values into a range of 0 to 1.

Clustering Approach

The number of clusters is determined using the Elbow Method to help produce representative groupings with clear differences between clusters. This approach measures cluster compactness using the Inertia metric for K-Means or Total Cost for K-Medoids. Both metrics measure the distance between data points and cluster centers (centroids or medoids), where lower values indicate tighter clusters.

Algorithm Implementation

Product clustering was performed using two main algorithms: K-Means and K-Medoids, with the optimal number of clusters (K) determined as 2 based on the Elbow Method. The K-means algorithm is performed in four stages, including initial centroid determination, data point placement into clusters, centroid position updating, and process termination when centroid changes are no longer significant or the maximum number of iterations has been reached (Li et al., 2024).

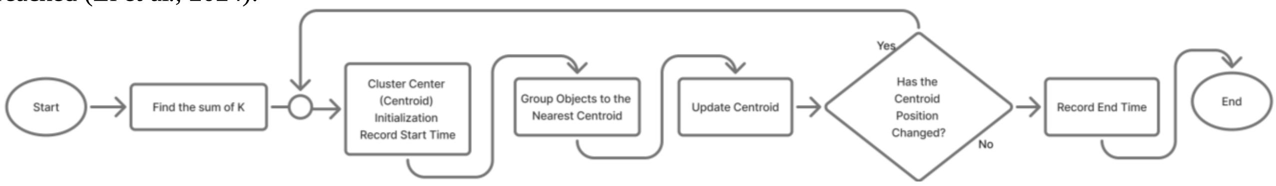

Fig 2. Flowchart K-Means

In comparison, K-Medoids uses medoids (the most central real data points) as cluster centers, making it more robust to outliers but computationally more complex.

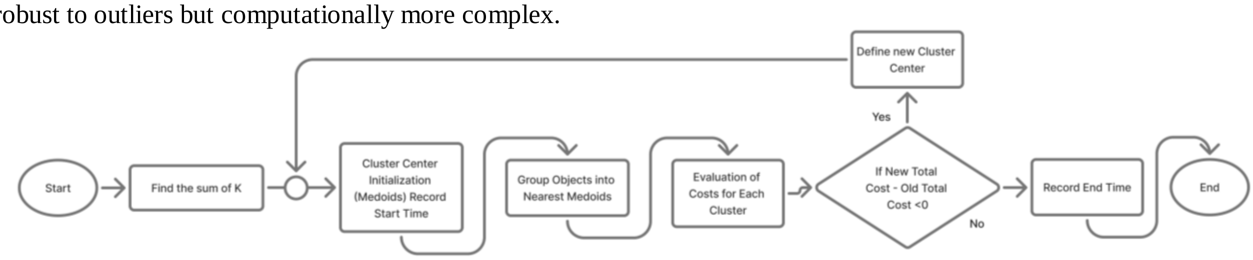

Fig 3. Flowchart K-Medoids

Evaluation

The evaluation stage is conducted to assess the quality of the clustering results produced by the K-Means and K-Medoids algorithms. The clustering results are evaluated using two popular metrics, namely the Davies Bouldin Index (DBI) and Silhouette Score, to ensure the quality of the clustering produced. These two metrics have been proven effective in measuring cluster validity based on the proximity and separation between data groups (Syahkur & Hartama, 2024). The Davies Bouldin Index (DBI) can be calculated using the following formula:

$$DBI = \frac{1}{k}\sum_{i=1}^{k} \max_{j \neq i} \left(\frac{\sigma_i+\sigma_j}{d(c_i,c_j)}\right) \qquad (1)$$

- σ_i : average distance between data and cluster center i
- $d(c_i, c_j)$: distance between cluster centers i and j

Meanwhile, the Silhouette Score can be calculated using the following formula:

$$s(i) = \frac{b(i)-a(i)}{\max(a(i),b(i))} \qquad (2)$$

- ai: is the average distance between i and other observations in a cluster

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 5, **Number** 2, November 2025
https://doi.org/10.47709/brilliance.v5i2.6992

- bi: the average distance between i and the observation in the nearest cluster

The calculation and analysis of the Davies-Bouldin Index (DBI) and Silhouette Score will be performed separately for the clustering results using the K-Means and K-Medoids algorithms. This aims to obtain a clear picture of the performance of each algorithm and to ensure that the evaluation is objective and does not influence each other.

Interpretation

After the entire process, from data collection, preprocessing, transformation, to determining the optimal number of clusters, is complete, the next step is to interpret the clustering results. The interpretation of the results aims to provide a deeper understanding of the structure and characteristics of each product group formed based on the clustering results.

RESULT

Data Preprocessing

At this stage, raw data was collected from the PRDECT-ID (Product Reviews Dataset for Emotions Classification Tasks – Indonesian) dataset uploaded to the Mendeley platform in 2022, which contained 5,401 entries covering the attributes Category, Product Name, Location, Price, Overall Rating, Number Sold, Total Review, Customer Rating, Customer Review, Sentiment, and Emotion. This preprocessing stage involved standardizing the writing in the Product Name and Category columns, mapping categories into 8 Main Category groups, removing empty and duplicate values, selecting categories with sufficient product entries, and handling outliers in the numeric attributes Price, Number Sold, and Total Review using the Interquartile Range (IQR) method.

After all stages were completed, the amount of data, which originally contained 5,401 entries, was reduced to 1,838 entries, but this amount was still sufficient for the next stage.

Data setelah cleaning (10 entri acak):

No	Category	Main Category	Product Name	Price	Customer Rating	Number Sold	Total Review
1507	mother and baby	Kesehatan	kodomo baby tisu basah anti bacterial 50 sheets	37400	2	5309	1636
1434	phones and tablets	Elektronik	samsung galaxy a12 4/128 gb garansi resmi sein - hitam	2114000	1	17400	4629
1484	mother and baby	Kesehatan	transpulmin baby balsam - 20gr	71390	4	4098	1925
1716	health	Kesehatan	tim ayam obat herbal komplit 12 macam	30000	3	7474	1311
998	muslim fashion	Fashion	hijab voal segiempat premium - emikoawa jilbab kerudung terbaru korea - capucino	28999	3	11700	2819
1190	men's fashion	Fashion	baju seragam pgri kemeja hem katun pria batik pgri dinas pns panjang - s	119900	3	525	259
576	sport	Otomotif	basic headband black / bando hitam pria wanita / aksesoris olahraga	4900	2	2806	491
332	animal care	Rumah	obat kutu hewan anjing & kucing bahan alami racoon / flea remover	79000	1	8243	3964
619	books	Hiburan	i saw the same dream again	83000	5	234	152
110	toys and hobbies	Hiburan	tenda bermain anak model castle (biru/ pink)	120000	1	1592	1085

Fig 4. Data After Cleaning

Data Transformation

Once clean data has been obtained, the next step is to transform the main numerical attributes so that they have a uniform scale. The attributes used are: Price, Customer Rating, Number Sold, and Total Review. Normalization is performed using the Min-Max Scaling method using the sklearn.preprocessing.MinMaxScaler library, which converts numerical values to a range of 0–1. Before transformation, the old "No" numbering column was deleted and replaced with new numbering to maintain the data order. After the normalization process was complete, the results were recombined with the descriptive columns (Main Category, Category, Product Name).

Optimal Clustering Approach

Before performing clustering with the K-Means and K-Medoids algorithms, optimization was carried out using the Elbow Method approach on two different algorithms, namely K-Means and K-Medoids, by calculating the inertia value (for K-Means) and total cost (for K-Medoids) in the range K = 1 to K = 10.

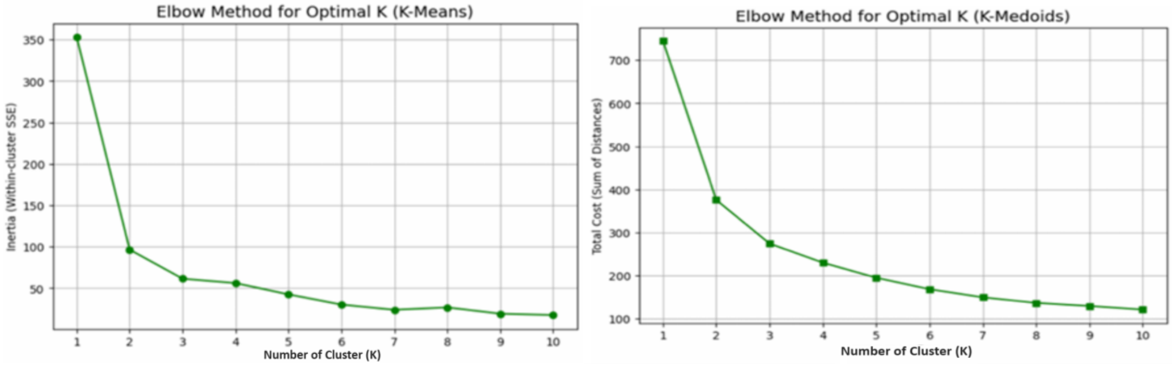

Fig 5. Graphical Visualization of K Values Using the Elbow Method

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

E-ISSN : 2807-9035
Volume 5, **Number** 2, November 2025
https://doi.org/10.47709/brilliance.v5i2.6992

The visualization results on both Elbow graphs show that the sharpest decline occurs at K = 2. Therefore, the optimal number of clusters used in the algorithm implementation is set at two clusters (K = 2).

Algorithm Implementation

The entire process of implementing the K-Means and K-Medoids algorithms was carried out using Python scripts, utilizing the sklearn.cluster.KMeans library for the implementation of K-Means, and sklearn.metrics.pairwise_distances for calculating the distance between data in K-Medoids. Each algorithm was run iteratively until it reached a convergent condition, which is when the position of the centroid or medoid no longer changed.

In the iterative process of K-Means implementation, after the calculation in the sixth iteration, the centroid position did not change compared to the previous iteration. This indicates that the clustering process has reached convergence, where all data has stabilized in its nearest cluster. The two final centroids that were formed represent each cluster, namely:

a. C1: [0.02354793 0.08096591 0.08210225 0.14972725]
b. C2: [0.04022589 0.82706374 0.06358168 0.12097744]

In the iterative process of implementing K-Medoids, the iteration stops at the fifth trial as the final medoid for each cluster. The process is stopped because the structure is already optimal and no additional iterations are needed.

Final Total Cost: 403.815028 Final Medoid (Unchanged):

a. C1 (Index 1336): [0.01685672 0. 0.06773229 0.11960605]
b. C2 (Index 1154): [0.01685672 1. 0.04401312 0.08579551]

Clustering Evaluation

After the clustering process is complete, the next step is to evaluate the quality of the grouping using quantitative metrics, namely the Davies-Bouldin Index (DBI) and Silhouette Score. The evaluation results can be seen in Table 1 below:

Table 1. DBI Evaluation Value and Silhouette Score

Algorithm	*Davies-Bouldin Index*	*Silhouette Score*
K-Means	0.5717	0.6012
K-Medoids	0.5870	0.5857

From the table above, it can be seen that the DBI value in K-Means indicates that the clusters formed have a fairly good separation between each other and the Silhouette Score value is in the good category. For the DBI value, K-Medoids also shows acceptable results and is able to reflect fairly stable clustering performance, as does the Silhouette Score, which also shows adequate results.

In addition to evaluating the structural quality of the clusters, this study also measures the efficiency of each algorithm through computation time. The results of the computation time comparison between the two algorithms can be seen in Table 2 below:

Table 2. Comparison of K-Means and K-Medoids Computation Time

Algorithm	Execution Time (seconds)	Description
K-Means	0.0947	Faster
K-Medoids	0.1622	More Complex Processes

Interpretation of Results

The clustering of Tokopedia products using K-Means and K-Medoids produced two main clusters (C1 and C2) based on price, category, rating, number of reviews, and number of sales. The difference in the number of cluster members reflects the difference in center determination, where K-Means uses centroids and K-Medoids uses medoids.

Table 3. Number of Products per Cluster from K-Means and K-Medoids

Method	Cluster	Number of Product
K-Means	C1	880
K-Means	C2	957
K-Medoids	C1	973
K-Medoids	C2	864

Descriptive analysis of numerical attributes (Price, Customer Rating, Number Sold, and Total Reviews) shows differences in characteristics between the two clusters. Cluster C1 contains products with higher average sales and reviews, representing popular products with strong commercial performance. However, the rating values in this cluster tend to be lower. Conversely, cluster C2 has higher ratings but lower sales and reviews, describing quality products that have not yet reached a wide market. Price variations are relatively balanced so that they are not a dominant factor in

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

E-ISSN : 2807-9035
Volume 5, **Number** 2, November 2025
https://doi.org/10.47709/brilliance.v5i2.6992

cluster division. An example of a product in C1 is the i8 mini touchpad wireless keyboard mouse with high sales and reviews, while the 12-ingredient herbal medicine product is included in C2 because it has high ratings despite low sales. These findings indicate that C1 is the cluster with high commercial performance, while C2 stands out in product quality. The recommended strategy is to maintain service quality in C1 and increase promotion and visibility in C2 to expand the market.

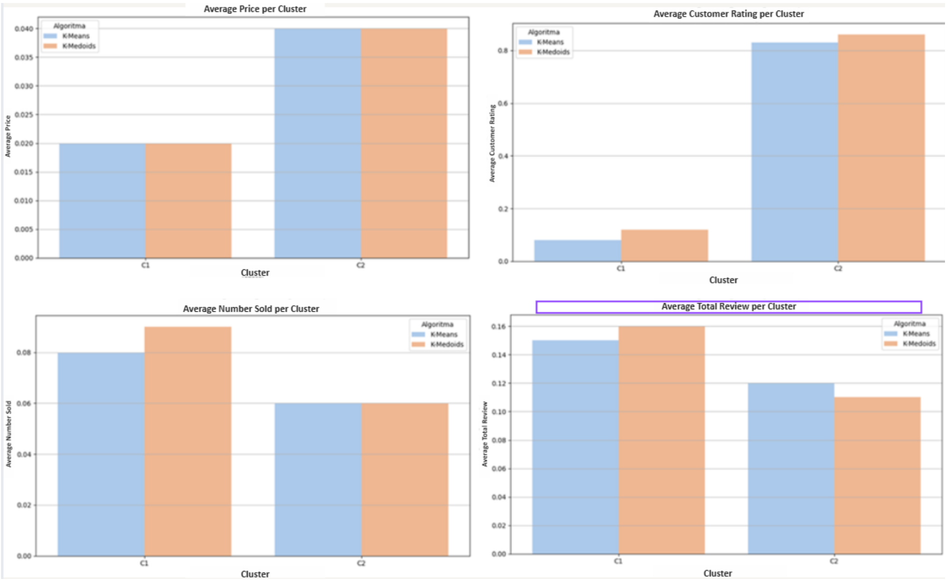

Fig 6. Analysis of the Characteristics of Each Cluster

Analysis of product category distribution shows differences in trends between each cluster. The Electronics and Home categories are relatively evenly distributed across both algorithms, while Fashion and Entertainment are more dominant in C2, indicating high-quality products but with low sales volumes. Conversely, the Health and Automotive categories are stronger in C1, representing high commercial performance. The General and Travel categories tend to be concentrated in C2 with low sales. These findings provide strategic direction: increase promotion and distribution in categories in C2, while maintaining service quality and expanding market reach in categories in C1. Overall, category distribution provides important insights for designing more targeted marketing strategies in line with established consumer segmentation.

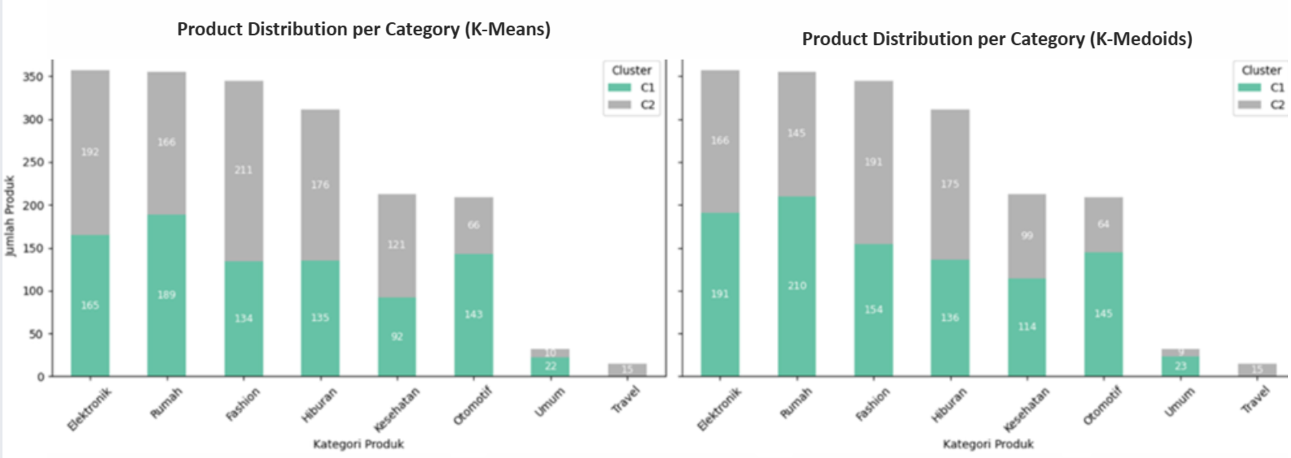

Fig 7. Product Distribution Comparison

DISCUSSION

The results of this study show that product clustering on Tokopedia using the K-Means and K-Medoids algorithms successfully grouped products based on numerical attributes such as price, rating, number of sales, and total reviews. The most significant finding was the difference in performance between the two algorithms. K-Means demonstrated superiority in forming a more focused and consistent cluster structure, producing clear cluster profiles. For example, one cluster contained products with high commercial performance (sales and reviews) and another cluster contained products with high quality (ratings). These differences in profiles provide nuanced and useful insights for e-commerce managers.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

E-ISSN : 2807-9035
Volume 5, **Number** 2, November 2025
https://doi.org/10.47709/brilliance.v5i2.6992

These findings are consistent with several previous studies that highlight the advantages of K-Means in terms of speed and efficiency for large datasets. However, this study also confirms the limitations of K-Means, which is theoretically sensitive to outliers. On the other hand, K-Medoids, which is known to be more resistant to outliers, shows slightly lower performance than K-Means in terms of cluster quality, although the results are still considered adequate. The evaluation results using the Davies-Bouldin Index (DBI) and Silhouette Score reinforce these findings, with K-Means producing a lower DBI value (0.5717) and a higher Silhouette Score (0.6012) compared to K-Medoids (DBI: 0.5870; Silhouette Score: 0.5857). This confirms that K-Means creates more compact and clearly separated clusters. Furthermore, a comparison of computation time also shows that K-Means is significantly faster (0.0947 seconds) than K-Medoids (0.1622 seconds), in line with existing literature on the efficiency of K-Means on large datasets.

One of the main limitations of this study is the use of secondary data from a specific dataset, which may not fully reflect the real-time dynamics of the Tokopedia e-commerce market. In addition, this study focuses only on four numerical attributes (Price, Customer Rating, Number Sold, Total Review), which limits the analysis and interpretation of cluster profiles. Another limitation is that the analysis does not include non-numerical attributes such as review sentiment or product features that could provide richer insights. For future research, it is recommended to use a more up-to-date dataset and include additional attributes to produce a more comprehensive product segmentation. It is also necessary to consider testing other algorithms or combinations of clustering methods for stronger result validation.

CONCLUSION

This study successfully applied clustering to Tokopedia product data using the K-Means and K-Medoids algorithms, grouping products based on the attributes of Price, Customer Rating, Number Sold, and Total Reviews. The results showed that the K-Means algorithm was superior in forming a directed and consistent cluster structure, where the clusters formed had clear performance profiles, for example, products with high commercial performance (sales and reviews) versus products with high quality (ratings). Evaluation using the Davies-Bouldin Index (DBI) and Silhouette Score metrics confirms that K-Means produces more compact and clearly separated clusters and is proven to be more efficient in terms of computation time compared to K-Medoids. The main contribution of this research is to provide a new analytical framework that enables e-commerce managers to gain comprehensive insights into product profiles, which can ultimately guide more effective marketing strategies and recommendation systems.

REFERENCES

Ainur Rahman, & Suroyo, H. (2021). Analisis Data Produk Elektronik Di E-Commerce Dengan Metode Algoritma K-Means Menggunakan Python. Journal of Advances in Information and Industrial Technology, 3(2), 11–18. https://doi.org/10.52435/jaiit.v3i2.158

Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J. M., & Perona, I. (2013). An extensive comparative study of cluster validity indices. *Pattern Recognition*, *46*(1), 243–256. https://doi.org/10.1016/j.patcog.2012.07.021

Arora, P., Deepali, & Varshney, S. (2016). Analysis of K-Means and K-Medoids Algorithm for Big Data. *Physics Procedia*, *78*(December 2015), 507–512. https://doi.org/10.1016/j.procs.2016.02.095

Gymnastiar, S., & Bahtiar, A. (2024). Penerapan Algorima K-Means Clustering Untuk Mengelompokan Data Kejadian Kekeringan Di Kabupaten Cirebon. JATI (Jurnal Mahasiswa Teknik Informatika), 8(2), 2325–2331. https://doi.org/10.36040/jati.v8i2.8948

Harjono, S. W., Utami, N. W., & Putri, I. G. A. P. D. (2023). Klasterisasi Tingkat Penjualan pada Startup Panak.id dengan Algoritma K-Means. Jurnal Ilmiah Teknologi Informasi Asia, 17(1), 55–66. https://doi.org/10.32815/jitika.v17i1.888

Hermawati, A., Jumini, S., Astuti, M., Ismail, F., & Rahim, R. (2020). Unsupervised Data Mining with K-Medoids Method in Mapping Areas of Student and Teacher Ratio in Indonesia. *TEM Journal*, *9*(4), 1614–1618. https://doi.org/10.18421/TEM94-37

Idrus, A., Tarihoran, N., Supriatna, U., Tohir, A., Suwarni, S., & Rahim, R. (2022). Distance Analysis Measuring for Clustering using K-Means and Davies Bouldin Index Algorithm. *TEM Journal*, *11*(4), 1871–1876. https://doi.org/10.18421/TEM114-55

Kulkanjanapiban, P., & Silwattananusarn, T. (2025). A Performance-Driven Exploration of Combining Topic Modeling and Machine Learning for Online Learning Data Analysis. *TEM Journal*, *14*(1), 511–527. https://doi.org/10.18421/TEM141-46

Li, S., Lim, C. Y., & Ang, S. L. (2024). An Analysis of Technostress Factors Among Teachers in Hunan, China Through Statistical Methods and K-means Clustering. *TEM Journal*, *13*(4), 3231–3240. https://doi.org/10.18421/TEM134-57

Lukáč, J., Kudlová, Z., Kopčáková, J., & Gallo, P. (2025). Impact of Socio-Economic Factors on Digital Literacy and Security. *TEM Journal*, *14*(1), 925–932. https://doi.org/10.18421/TEM141-81

Purba, D. S., Dwi Permatasari, P., Tanjung, N., Rahayu, P., Fitriani, R., Wulandari, S., Universitas,), Negeri, I., Utara, S., Muslim, U., & Al Washliyah, N. (2025). Analisis Perkembangan Ekonomi Digital Dalam Meningkatkan

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

E-ISSN : 2807-9035
Volume 5, **Number** 2, November 2025
https://doi.org/10.47709/brilliance.v5i2.6992

Pertumbuhan Ekonomi Di Indonesia. Jurnal Masharif Al-Syariah: Jurnal Ekonomi Dan Perbankan Syariah, 10(1), 126–139.

Rhomadhona, H., Kusrini, W., Aprianti, W., & Permadi, J. (2025). Implementation of K-Means Clustering for Social Assistance Recipients with Silhouette Score Evaluation. *Brilliance: Research of Artificial Intelligence*, *5*(1), 136–143. https://doi.org/10.47709/brilliance.v5i1.5900

Riyahi, M., & Martín, A. G. (2025). Optimizing capacity expansion modeling with a novel hierarchical clustering and systematic elbow method: A case study on power and storage units in Spain. *Energy*, *323*(March). https://doi.org/10.1016/j.energy.2025.135788

Saputri, F. W., & Arianto, D. B. (2023). Perbandingan Performa Algoritma K-Means, K- Medoids, Dan Dbscan Dalam Penggerombolan Provinsi Di Indonesia Berdasarkan Indikator Kesejahteraan Masyarakat. Jurnal Teknologi Informasi, 17(2), 138–151.

Sari, V. K., & Nasution, M. I. P. (2024). Dampak E-commerce Terhadap Perkembangan Digital. Jurnal Akademik Ekonomi Dan Manajemen, 1(4), 18–24.

Syahkur, M. R., & Hartama, D. (2024). Evaluasi Jumlah Cluster pada Algoritma K-Means ++ Menggunakan Silhouette dan Elbow dengan Validasi Nilai DBI dalam Mengelompokkan Gizi Balita. Jurnal Sains Dan Teknologi, 13(3), 487–496.

Umagapi, I. T., & Umaternate, B. (2023). Uji Kinerja K-Means Clustering Menggunakan Davies-Bouldin Index Pada Pengelompokan Data Prestasi Siswa. Prosiding Sisfotek, 303–308.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.