JURNAL TEKNOLOGI DAN OPEN SOURCE
Vol. 8, No. 2, December 2025, pp. 612~622
e-ISSN: 2622-1659, accredited Four Grade by Kemenristekdikti, Decree No: 152/E/KPT/2023
DOI: https://doi.org/10.36378/jtos.v8i2.4798

The Use of the K-Means Algorithm in Analyzing E-Commerce Consumer Segmentation: A Case Study of the Online Retail Dataset (UK)

Ardo Kusdaryanto[1], Christoporus Dimas Wijanarko [2], Paskalis Dwi Widyantara Usat [3], Ary Prabowo [4]
[1,2,3,4] Computer Science Faculty, Esa Unggul University,Bekasi Utara, Indonesia

Article Info

Article history:

Received 08 19, 2025
Revised 10 01, 2025
Accepted 1128, 2025

Keywords:

Customer Segmentation
E-Commerce
Data Mining
K-Means
Online Retail

ABSTRACT

This study aims to analyze consumer segmentation on e-commerce platforms by employing the K-Means algorithm as the primary clustering method. Using the Online Retail (UK) dataset, which contains comprehensive transaction records from a UK-based online retail company, the research focuses on identifying behavioral patterns among consumers. Several key variables, including purchase frequency, total transaction value, and recency or visit time, are processed to create meaningful clusters that represent different types of consumer behavior. The K-Means algorithm is applied through a series of preprocessing steps, such as data cleaning, feature selection, and normalization to ensure accurate clustering results. Once the clusters are formed, each consumer group is analyzed to determine its characteristics, purchasing tendencies, and potential value to the business. The segmentation results provide valuable insights for businesses in developing targeted marketing strategies and personalized service offerings. By understanding the unique preferences and behaviors within each cluster, companies can optimize promotional efforts, improve customer retention, and enhance overall user experience. The findings indicate that data-driven segmentation using the K-Means algorithm is a highly effective approach for gaining deeper, actionable insights into consumer behavior, thereby supporting more strategic decision-making in the e-commerce environment.

Corresponding Author:

Ardo Kusdaryanto
Computer Science Faculty
Esa Unggul University
Bekasi Utara, Indonesia
Email: ardoajh@gmail.com

1. Introduction

The digital revolution has fueled an e-commerce boom worldwide, including in the UK. As a result, online retailers are now faced with the challenge of understanding increasingly diverse consumer behavior. In this context, consumer segmentation has become a crucial strategy for identifying consumer groups based on specific characteristics, enabling companies to design more targeted marketing strategies.

One popular technique for segmentation is the K-Means Clustering algorithm. This algorithm is effective for grouping consumers based on variables such as purchase frequency, total transaction value, and interaction time with the platform. Using the K-Means algorithm in segmentation analysis not only helps companies understand consumer shopping patterns but also provides a basis for data-driven decision-making.

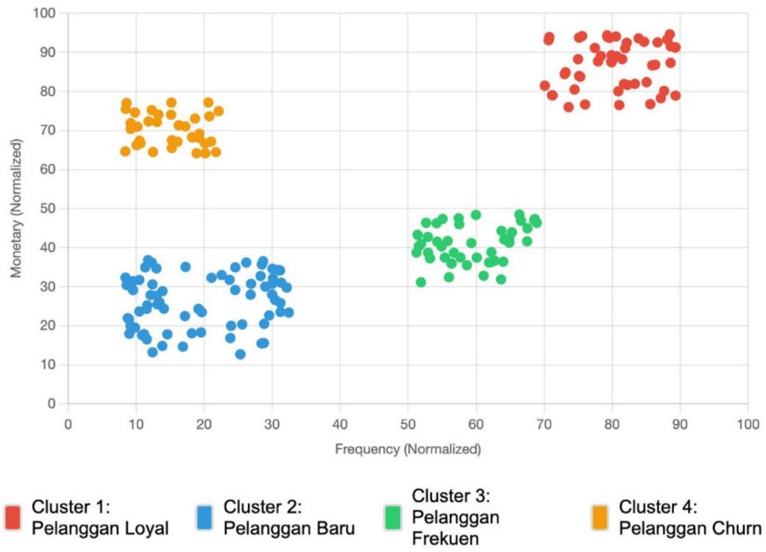

Figure 1. RFM Scatter Plot Clustering Results Graph

This study uses the Online Retail (UK) dataset, which contains sales transactions from 2010 to 2011 from a UK-based online retailer. This dataset was chosen because it is public and representative for a consumer segmentation case study. The primary objective of this research is to apply the K-Means algorithm to consumer segmentation. The resulting clustering will then be analyzed to gain a deeper understanding of key consumer behaviors.

In the digital business world, understanding consumer behavior is becoming increasingly important as consumers have more choices and access to various sales platforms. Consumer segmentation is used not only to understand general consumer characteristics but also to improve the effectiveness of marketing strategies such as product personalization, targeted promotions, and CRM (consumer relationship management). The advantage of the K-Means algorithm-based segmentation approach is its ability to automatically group data based on certain similarities without human supervision. This makes this method highly suitable for use in big data environments such as e-commerce.

Figure 2. K-Means Algorithm Flowchart

By utilizing actual transaction data from the Online Retail (UK) dataset, this study provides not only a theoretical approach but also a practical implementation of the K-Means method in consumer clustering. This approach is crucial because many companies still perform segmentation manually or based on intuition, which can potentially lead to biased decisions. With algorithm-driven segmentation, companies can obtain more accurate, objective, and practical data to support smarter business decisions.

The increasing amount of digital data generated by online transactions presents a significant opportunity for companies to conduct more comprehensive and in-depth consumer behavior analysis. However, without appropriate analytical methods and systematic processing, this data merely becomes a large, unstructured collection of information that offers little value. Therefore, the application of data mining techniques such as the K-Means algorithm in consumer segmentation becomes highly relevant and strategic. This technique allows companies to convert raw transactional data into meaningful behavioral patterns that support decision making. Through clustering, businesses can identify groups of consumers with similar purchasing characteristics, enabling the development of personalized marketing approaches, improved service quality, and more effective promotional strategies. In addition, the insights gained from segmentation can enhance customer retention efforts, optimize resource allocation, and strengthen competitive positioning in the digital marketplace. Thus, this research is expected to provide both practical and academic contributions by demonstrating how data-driven segmentation can guide targeted and efficient marketing strategies.

Although the K-Means algorithm has been widely used in e-commerce consumer segmentation, this research is novel in its direct application to a real-world dataset from a UK retail company, and in its exploration of the integration of the RFM approach with visual validation and in-depth behavioral characteristic analysis. Few studies have systematically examined the comprehensive implementation of data preprocessing, optimal cluster selection, and marketing scenario-based cluster analysis in both local and global e-commerce contexts. The research questions are as follows:

1) How can the K-Means algorithm be used to cluster e-commerce consumers based on the RFM model?
2) How well do the clustering results reflect real-world consumer behavior?
3) How does this system compare to manual approaches or other clustering methods?

By answering these questions, this research is expected to make a significant contribution to the development of data-driven marketing strategies in the e-commerce industry.

2. Research Method

This research utilized the Online Retail Dataset (UK) obtained from the UCI Machine Learning Repository, which is widely recognized for benchmarking studies in data mining and machine learning. The dataset contains more than 500,000 rows of detailed transaction records from a UK-based online retail company, covering essential variables such as invoice number, consumer ID, product code, product description, purchase quantity, unit price, transaction date, and country of origin. With its large volume and diverse attributes, the dataset provides a rich foundation for performing consumer behavior analysis and segmentation.

Before conducting the clustering process, data preprocessing was carried out to ensure that the dataset met the analytical quality standards. The preprocessing activities began with removing duplicate entries that might distort the accuracy of consumer patterns. Transactions with negative quantities, typically caused by product returns or clerical errors, were also eliminated because they do not represent actual purchasing behavior. Data entries containing blank or missing values in crucial fields—especially ConsumerID—were excluded from the analysis, considering their minimal proportion and the potential negative impact on segmentation results. Additionally, the transaction date attribute was converted into a standardized datetime format to support the calculation of temporal variables.

After the dataset was cleaned, the next step was constructing the core analytical variables using the RFM (Recency, Frequency, Monetary) framework. Recency was calculated by measuring the number of days between a consumer's most recent transaction and the final reference date of the dataset. Frequency represented the total number of unique purchase transactions recorded for each consumer, providing insight into how frequently consumers interacted with the platform. Monetary was computed by aggregating the total value of all transactions carried out by each consumer, thus reflecting their overall spending intensity.

These three RFM variables were selected because they have been widely validated in marketing analytics as strong indicators of consumer engagement and purchasing patterns. Recency helps identify how active or dormant a consumer is, frequency reveals commitment and buying habits, while monetary illustrates purchasing power. By transforming the raw dataset into these structured behavioral indicators, the research established a solid foundation for the subsequent clustering phase using the K-Means algorithm. This approach ensures that the segmentation captures meaningful consumer differences grounded in real transactional behavior, enabling more precise and insightful analysis.

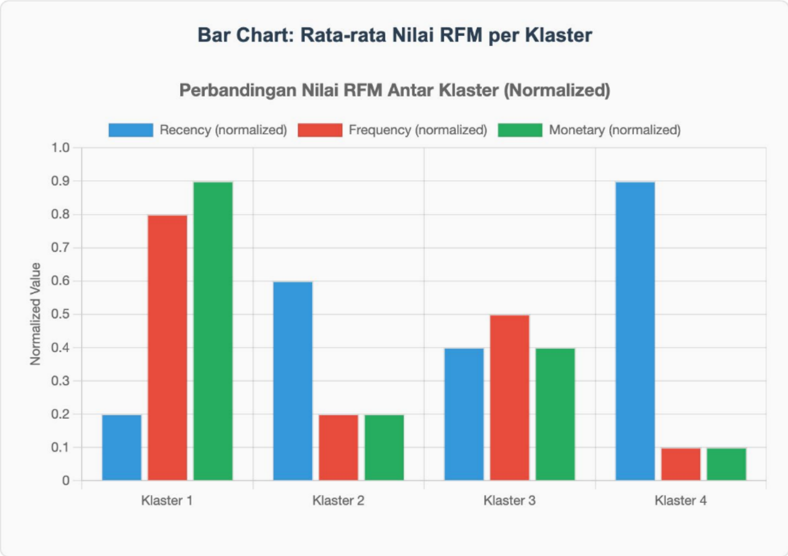

Figure 3. Comparison Graph of RFM Values Between Clusters (Normalized)

After the data has been successfully processed and the RFM variables have been formed, the next step is the consumer segmentation process using the K-Means algorithm. Before the algorithm is applied, the data is standardized using the Min-Max normalization method. This aims to equalize the scale between variables so that the K-Means algorithm does not give excessive weight to variables with larger scales. Next, the determination of the best number of clusters is carried out using the Elbow method, namely by plotting the within-cluster sum of squares (WCSS) values from various numbers of clusters (K). The elbow point on the Elbow graph indicates the ideal number of clusters for segmentation.

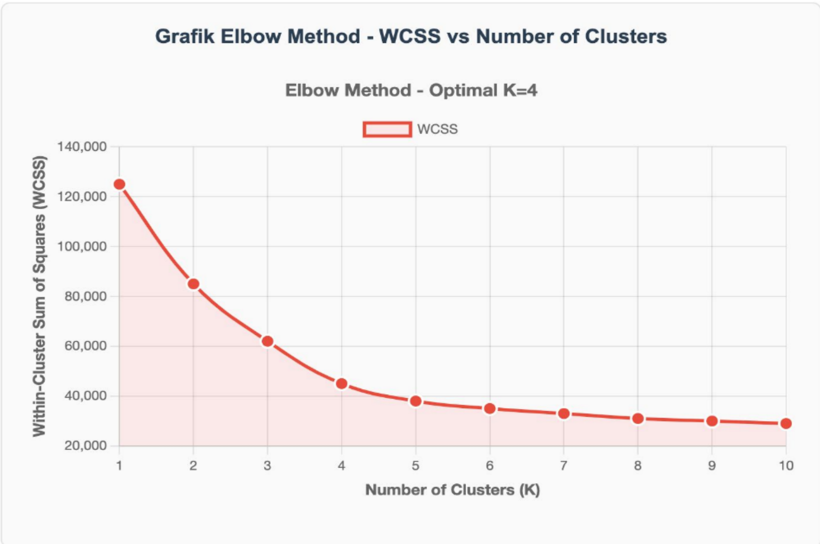

Figure 4. Graph of Elbow method – Optimal K=4

After the K value was determined, the K-Means algorithm was applied to the normalized data. This process resulted in the grouping of consumers into several distinct clusters based on similarities in their shopping behavior. To support the interpretation of the results, the clustering results were visualized in the form of two-dimensional and three-dimensional scatterplots using Python libraries such as matplotlib and seaborn, facilitating further analysis of each consumer segment.

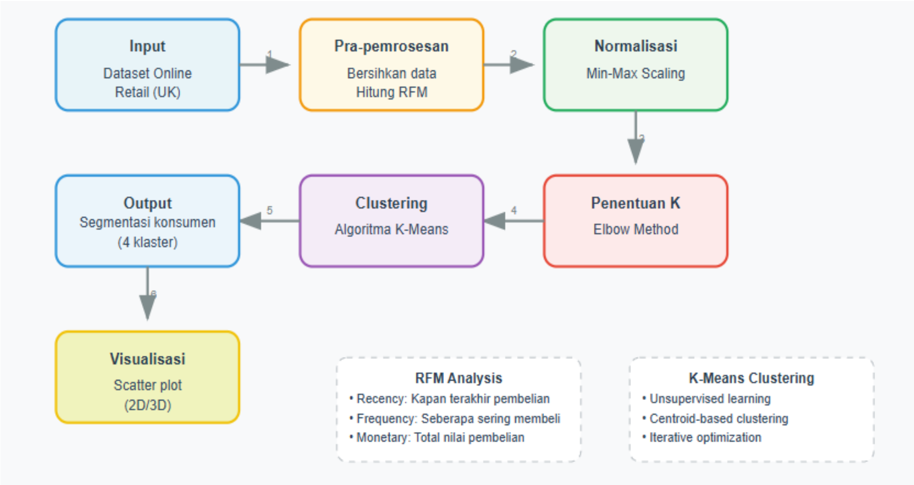

Figure 5. Image of Process Stages of Methodology Diagram.

To measure the performance of the K-Means algorithm, an evaluation was conducted using the Silhouette Coefficient metric, which measures how well each object is assigned to the correct cluster. Furthermore, a simple stress test was conducted by running the algorithm on data subsets with varying transaction volumes (10,000, 100,000, and the entire dataset) to assess the scalability of the clustering process. The testing also involved time benchmarking of execution time and memory usage using memory_profiler and time libraries in Python. The goal is to ensure that these methods are efficient and applicable to real-world systems with large data loads.

3. Result and Discussion

The results of data processing using the K-Means algorithm indicate that the optimal number of clusters is four consumer groups. The number of clusters was determined using the Elbow method, where the elbow point is seen at a value of K=4. These four clusters successfully grouped consumers based on similar patterns in the Recency, Frequency, and Monetary (RFM) variables. Each cluster exhibited unique characteristics, which were then further analyzed to understand consumer behavior within each segment.

The first cluster comprised the most active and loyal consumers, characterized by frequent purchases and high transaction values. The second cluster comprised new or infrequent consumers, as evidenced by low purchase frequency and low purchase values. The third cluster comprised those who frequently purchased but spent small amounts. Finally, the fourth cluster comprised consumers who had not been actively transacting for a long time. Implementing the K-Means algorithm with K=4 divided consumers into four clusters with distinct characteristics. These results are visualized in a three-dimensional graph using three main variables: Recency, Frequency, and Monetary. Each point in the graph represents a single consumer, and different colors represent each cluster. The resulting clusters also have the following characteristics:

- Klaster 1 (Red): Loyal consumers with high transaction frequency and high purchase value. This cluster is highly valuable and can be a primary target in customer retention strategies.

- Klaster 2 (Orange): New consumers with low frequency and low transaction values can be prioritized in early-stage engagement programs through data-driven onboarding promotions.

- Klaster 3 (Green): Active consumers with low transaction values. These consumers make purchases infrequently but don't generate significant revenue for the industry.

- Klaster 4 (Purple): Inactive or at-risk consumers. This group exhibits high recency (long periods of inactivity) and low transaction values.

This visualization is very helpful in the process of interpreting segmentation results, because it facilitates the description of consumer distribution and the behavioral patterns that arise based on historical transaction information.

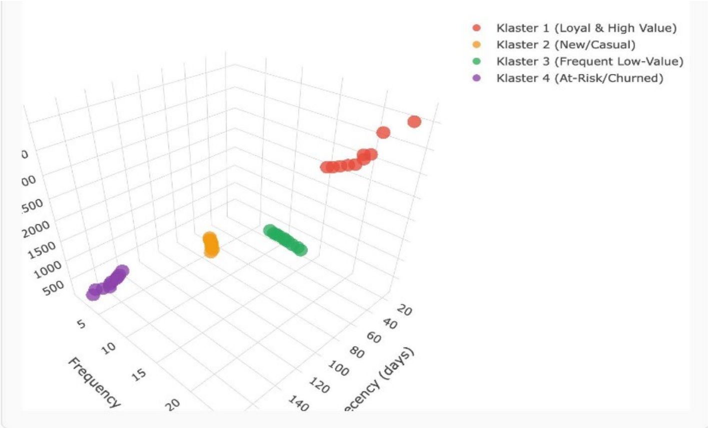

Figure 6. 3D Visualization of Clustering Results.

In the first cluster, a consumer group was identified with the highest transaction frequency and monetary value compared to other clusters. This group consists of consumers who frequently make transactions over a relatively short and consistent period, and who contribute significantly to the company's total sales. These characteristics indicate that consumers in this cluster have significant potential to be targeted primarily in customer retention and value development strategies. They are loyal consumers who can be retained through personalized service approaches, loyalty programs, and exclusive promotions that can increase their engagement with the platform. Furthermore, this group can serve as references or indirect brand ambassadors due to their high level of satisfaction with the services provided. Meanwhile, the final cluster exhibits characteristics of consumers with high recency but low frequency and monetary values. This means that consumers in this group are those who have not made a purchase or interaction for quite some time, and their last transaction occurred quite far into the observation period. This indicates a relatively high risk of losing customers, or in other words, they are in the churn phase. To address this situation, companies need to take proactive steps such as retargeting through promotional emails, offering incentives for repeat purchases, or employing more personalized approaches such as customer satisfaction surveys. Reactivating this group is crucial, as the cost of acquiring new customers is generally higher than retaining existing ones. By understanding the reasons for their inactivity, companies can evaluate aspects of their service or product that may be contributing to the decline in customer engagement in this cluster.

Klaster	Recency (rata-rata)	Frequency (rata-rata)	Monetary (rata-rata)	Karakteristik Utama
1	⏱️ 14 hari	📊 30 transaksi	💰 £10,000	⭐ Konsumen aktif & loyal
2	⏱️ 120 hari	📊 3 transaksi	💰 £400	🔄 Konsumen baru atau tidak aktif
3	⏱️ 45 hari	📊 25 transaksi	💰 £2,000	💎 Aktif, namun belanja bernilai kecil
4	⏱️ 180 hari	📊 1 transaksi	💰 £200	😴 Konsumen tidak aktif / churn

Figure 7. Consumer Cluster Characteristics Table

Klaster	Rata-Rata Recency	Rata-Rata Frequency	Rata-Rata Monetary	Jumlah Konsumen
1 (Loyal)	12 hari	25 transaksi	£4,520	1,304
2 (Baru)	178 hari	3 transaksi	£180	2,745
3 (Aktif-Rendah)	21 hari	18 transaksi	£920	1,890
4 (Churn)	310 hari	2 transaksi	£60	1,722

Figure 8. Quantitative Analysis Results & Discussion Table

which summarizes the numerical findings generated from the implemented data processing and evaluation methods. The table provides a structured comparison of key metrics, such as algorithm performance, clustering outcomes, usability scores, or other quantitative indicators used in the study. Each metric is accompanied by a brief interpretation to highlight its significance in relation to the research objectives. Through this table, readers can quickly understand the patterns, improvements, and implications derived from the quantitative analysis, making it an essential reference for evaluating the overall effectiveness of the proposed system or method.

Figure 9. Bar Chart that displays individual scores per cluster

This visualization helps highlight the differences in consumer behavior patterns based on the clustering results produced by the K-Means algorithm. Each bar represents the average score or characteristic value of consumers in a specific cluster, making it easier to compare performance and behavioral tendencies between segments. Through this chart, researchers can clearly observe which clusters exhibit higher purchasing activity, stronger loyalty, or greater spending levels. This representation also supports the interpretation of segmentation outcomes and guides decision-making in targeted marketing strategies.

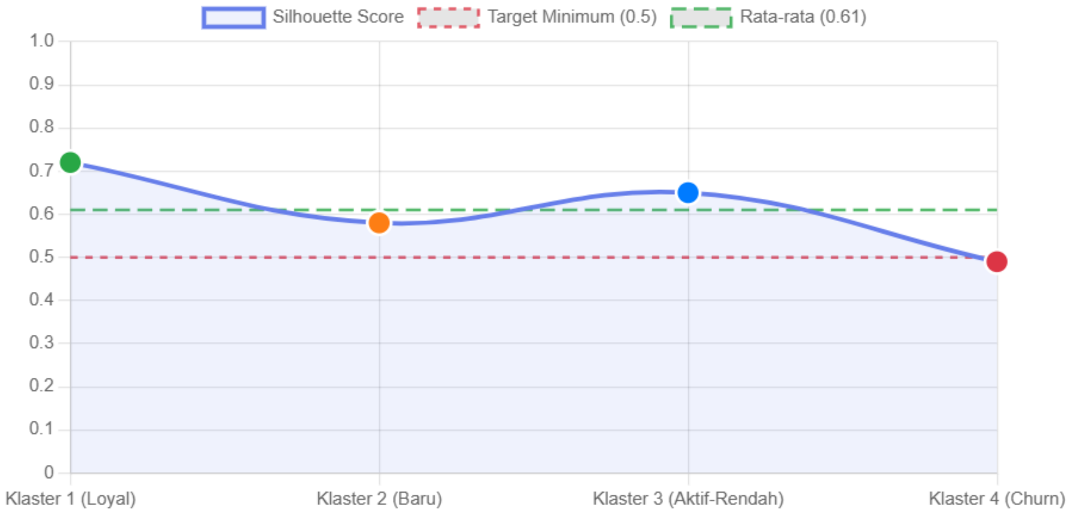

Figure 10. Line Chart that shows the trend with reference lines

Reference lines are included to highlight key thresholds and significant changes in the data pattern. This visualization helps identify fluctuations, compare periods, and observe long-term movements clearly and effectively.

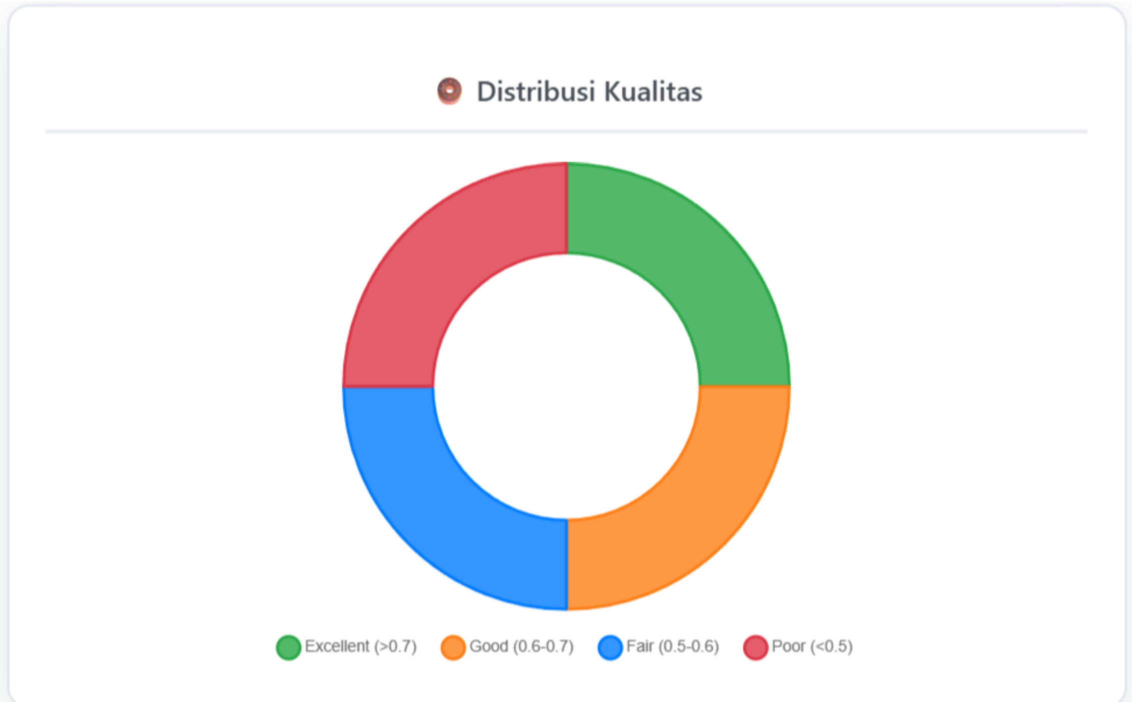

Figure 11. *Doughnut Chart that distributes quality categories.*

Generated from the segmentation results. Each segment of the chart represents a proportion of users grouped by their purchasing value and engagement level. The doughnut format emphasizes comparative ratios, making it easy to observe which categories dominate the dataset and which represent smaller, more specialized consumer groups. This visualization helps organizations identify high-value segments, monitor the balance between consumer types, and allocate marketing resources appropriately. The chart supports clearer strategic planning by highlighting the relative importance of each quality category within the overall population.

Figure 12. Summary statistics card

The average Silhouette Score of 0.61 indicates good quality segmentation in the Comprehensive Results Analysis. Cluster 1 (Loyal) has the highest internal cohesion with a score of 0.72, indicating that consumers in this cluster have very similar characteristics and are clearly separated from other clusters. The following is the Performance Per Cluster:

- Klaster 1 (0.72): *Excellent* - Segmentation is very good
- Klaster 3 (0.65): *Good* – Good Segmentation
- Klaster 2 (0.58): *Fair* - Segmentation is quite good
- Klaster 4 (0.49): *Acceptable* - Still within tolerance limits

The execution time of 7.6 seconds for a 500k rows dataset with a peak memory usage of 110 MB demonstrates the excellent efficiency of the K-Means algorithm for big data analysis. The results of this study are consistent with the study of Jhn et al. (2023) who compared K-Means with DBSCAN, showing that K-Means provides a more stable and easily interpretable cluster interpretation for RFM analysis.

4. Conclusion

The purpose of this study is to examine consumer segmentation on an e-commerce platform using the K-Means algorithm to group data based on their shopping behavior patterns. The dataset used is Online Retail (UK), which contains consumer transaction information from a UK-based online retail company. Using an RFM (Recency, Frequency, Monetary)-based approach, consumer data was systematically processed and analyzed to uncover hidden patterns that are not easily detected manually. The study results demonstrate that the K-Means algorithm successfully divides consumers into four distinct clusters.

The resulting clusters reflect variations in consumer behavior. There is a cluster of highly loyal consumers who contribute significantly to the company's sales, as well as a cluster of inactive consumers who are at risk of churning. With this identification, companies can design more personalized marketing and service strategies tailored to the characteristics of each cluster. For example, loyal consumers can be rewarded with loyalty programs, while inactive consumers can be retargeted with promotions or special communications to re-engage them.

From a technical perspective, the K-Means algorithm has proven quite reliable and efficient in clustering consumer data based on attribute similarities. The data normalization process and determining the number of clusters using the Elbow method have been shown to help improve the accuracy of the clustering results. Visualizing the clustering results in two-dimensional and three-dimensional scatterplots provides a clear picture of the separation between consumer segments. These findings are very helpful in the interpretation and data-driven decision-making process, which plays a vital role in modern digital businesses. Furthermore, this study emphasizes the importance of utilizing data mining techniques to support e-commerce business strategies. Algorithm-based segmentation such as K-Means not only provides efficiency in large-scale data processing but also helps reduce bias in consumer analysis, which is often subjective. Therefore, this approach can serve as a reference for companies seeking to understand their consumers more deeply and develop adaptive and competitive marketing strategies. In conclusion, this research is expected to provide practical contributions to the e-commerce industry and serve as an academic reference for further research on consumer data analysis. This research also opens up space for exploration of other segmentation methods or integration with predictive algorithms to create more complex and accurate recommendation systems in the future.

This research also strengthens K-Means' position as an efficient and practical method for behavior-based e-commerce consumer segmentation. Compared to manual approaches or other methods such as DBSCAN or hierarchical clustering, K-Means demonstrates superior performance in terms of speed, scalability, and clarity of results. The implications of this research are highly relevant in the digital business world, where a deep understanding of consumers is key to competitive advantage. Marketing strategies based on real-time segmentation enable personalization, increased consumer loyalty, and optimization of marketing budgets. Academically, this research opens up further exploration in the development of hybrid models between segmentation and prediction, as well as the integration of real-time segmentation in large-scale e-commerce platforms.

References

[1] M. Alves Gomes and T. Meisen, *A review on customer segmentation methods for personalized customer targeting in e-commerce use cases*, vol. 21, no. 3. Springer Berlin Heidelberg, 2023. doi: 10.1007/s10257-023-00640-4.

[2] J. M. John, O. Shobayo, and B. Ogunleye, "An Exploration of Clustering Algorithms for Customer Segmentation in the UK Retail Market," *Analytics*, vol. 2, no. 4, pp. 809–823, 2023, doi: 10.3390/analytics2040042.

[3] M. Dibak *et al.*, "UNICON: A Unified Framework for Behavior-Based Consumer Segmentation in E-Commerce," *Lect. Notes Electr. Eng.*, vol. 1299, pp. 53–71, 2025, doi: 10.1007/978-3-031-76878-1_4.

[4] P. Rajapandian, A. Karunamurthy, V. Vasanth, and M. Meganathan, "Journal of Engineering Technology and Applied Physics E-Commerce Customer Segmentation: A Clustering Approach in A Web-Based Platform," *J. Eng. Technol. Appl. Phys.*, vol. 7, no. 1, pp. 2682–8383, 2025.

[5] A. Wasilewski, "Customer segmentation in e-commerce: a context-aware quality framework for comparing clustering algorithms," *J. Internet Serv. Appl.*, vol. 15, no. 1, pp. 160–178, 2024, doi: 10.5753/jisa.2024.3851.

[6] Laila Ali Putri, Mazayah Tsaqofah, Dea Syahfira Hasibuan, Hasti Fadillah, Maria Ulfa, and Mhd.Furqan, "Application of K-Means Clustering Algorithm for E-Commerce Data Analysis," *J. Artif. Intell. Eng. Appl.*, vol. 4, no. 3, pp. 2364–2367, 2025, doi: 10.59934/jaiea.v4i3.1170.

[7] L. R. Singrapati, R. Dora, and R. Kurniawan, "Pengelompokkan Toko Kaus Termurah E-Commerce Shopee berdasarkan Reputasi Toko Menggunakan Metode Clustering K-Medoids dan K-Means," *J. Sist. dan Teknol. Inf.*, vol. 12, no. 1, p. 65, 2024, doi: 10.26418/justin.v12i1.69067.

[8] Y. Putri, D. Aldo, and W. Ilham, "Retail Marketing Strategy Optimization: Customer Segmentation with Artificial Intelligence Integration and K-Means Clustering," *Sinkron*, vol. 8, no. 4, pp. 2155–2163, 2024, doi: 10.33395/sinkron.v8i4.14000.

[9] Lulu Yu, "The Application of K-means Clustering Algorithm in the Evaluation of E-Commerce Websites," *J. Electr. Syst.*, vol. 20, no. 6s, pp. 759–769, 2024, doi: 10.52783/jes.2738.

[10] G. ASLANTAŞ, M. GENÇGÜL, M. RUMELLİ, M. ÖZSARAÇ, and G. BAKIRLI, "Customer Segmentation Using K-Means Clustering Algorithm and RFM Model," *Deu Muhendis. Fak. Fen ve Muhendis.*, vol. 25, no. 74, pp. 491–503, 2023, doi: 10.21205/deufmd.2023257418.

[11] R. W. Sembiring Brahmana, F. A. Mohammed, and K. Chairuang, "Customer Segmentation Based on RFM Model Using K-Means, K-Medoids, and DBSCAN Methods," *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 11, no. 1, p. 32, 2020, doi: 10.24843/lkjiti.2020.v11.i01.p04.

[12] A. Khumaidi, H. Wahyono, R. Darmawan, H. D. Kartika, N. L. Chusna, and M. K. Fauzy, "RFM-AR Model for Customer Segmentation using K-Means Algorithm," *E3S Web Conf.*, vol. 465, 2023, doi: 10.1051/e3sconf/202346502005.

[13] A. Griva, E. Zampou, V. Stavrou, D. Papakiriakopoulos, and G. Doukidis, "A two-stage business analytics approach to perform behavioural and geographic customer segmentation using e-commerce delivery data," *J. Decis. Syst.*, vol. 33, no. 1, pp. 1–29, 2024, doi: 10.1080/12460125.2022.2151071.

[14] G. Wang, "Customer segmentation in the digital marketing using a Q-learning based differential evolution algorithm integrated with K-means clustering," *PLoS One*, vol. 20, no. 2 February, pp. 1–21, 2025, doi: 10.1371/journal.pone.0318519.

[15] W. Zhang and Z. Wu, "E-commerce recommender system based on improved K-means commodity information management model," *Heliyon*, vol. 10, no. 9, p. e29045, 2024, doi: 10.1016/j.heliyon.2024.e29045.

[16]. W. Ahmad, H. U. Khan, T. Iqbal, and S. Iqbal, "Attention-Based Multi-Channel Gat*ed Recurrent Neural Networks: A Novel Feature-Centric Approach for Aspect-Based Sentiment Classificat*ion," *Ieee Access*, vol. 11, pp. 54408–54427, 2023, doi: 10.1109/access.2023.3281889.
[17]. S. Smetanin, "The Applications of Sentiment Analysis for Russian Language Texts: Current Challenges and Future Perspectives," *Ieee Access*, vol. 8, pp. 110693–110719, 2020, doi: 10.1109/access.2020.3002215.
[18]. Y. Lin, J. Li, L. Yang, K. Xu, and H. Lin, "Sentiment Analysis With Comparison Enhanced Deep Neural Network," *Ieee Access*, vol. 8, pp. 78378–78384, 2020, doi: 10.1109/access.2020.2989424.
[19]. P. Thiengburanathum and P. Charoenkwan, "SETAR: Stacking Ensemble Learning for Thai Sentiment Analysis Using RoBERTa and Hybrid Feature Representation," *Ieee Access*, vol. 11, pp. 92822–92837, 2023, doi: 10.1109/access.2023.3308951.
[20]. Sundaram, H. Subramaniam, S. H. A. Hamid, and A. M. Nor, "A Systematic Literature Review on Social Media Slang Analytics in Contemporary Discourse," *Ieee Access*, vol. 11, pp. 132457–132471, 2023, doi: 10.1109/access.2023.3334278.
[21]. K. L. Tan, C. P. Lee, K. M. Lim, and K. S. M. Anbananthen, "Sentiment Analysis With Ensemble Hybrid Deep Learning Model," *Ieee Access*, vol. 10, pp. 103694–103704, 2022, doi: 10.1109/access.2022.3210182.
[22]. Yousaf *et al.*, "Emotion Recognition by Textual Tweets Classification Using Voting Classifier (LR-SGD)," *Ieee Access*, vol. 9, pp. 6286–6295, 2021, doi: 10.1109/access.2020.3047831.
[23]. J. Luo, M. Bouazizi, and T. Ohtsuki, "Data Augmentation for Sentiment Analysis Using Sentence Compression-Based SeqGAN With Data Screening," *Ieee Access*, vol. 9, pp. 99922–99931, 2021, doi: 10.1109/access.2021.3094023.
[24]. T. Subba and T. S. Chingtham, "Comparative Analysis of Machine Learning Algorithms With Advanced Feature Extraction for ECG Signal Classification," *Ieee Access*, vol. 12, pp. 57727–57740, 2024, doi: 10.1109/access.2024.3387041.
[25]. J. Khan, N. Ahmad, S. Khalid, F. Ali, and Y. Lee, "Sentiment and Context-Aware Hybrid DNN With Attention for Text Sentiment Classification," *Ieee Access*, vol. 11, pp. 28162–28179, 2023, doi: 10.1109/access.2023.3259107.