E-ISSN: 2774-5694 P-ISSN: 2776-7388

International Journal of Management Science and Information Technology
IJMSIT

Volume 6 (1), January-June 2026, 765-770
DOI: https://doi.org/10.35870/ijmsit.v6i1.7213

Consumer Segmentation With K-Means at Lucky Shop Tanjungbalai

Reza Ahmad Fauzi [1*], Masitah Handayani [2], Parini [3]
[1*,2,3] Information Systems Study Program, Faculty of Computer Science, Universitas Royal, Asahan Regency, North Sumatra, Indonesia

Email: rezaahmadfauzi343@gmail.com [1*], bungafairuz8212@gmail.com [2], parini.royal@gmail.com [3]

Article history:
Received May 15, 2026
Revised May 30, 2026
Accepted June 4, 2026

Abstract

Consumer segmentation is an important strategy for improving marketing effectiveness and inventory management in retail businesses. Lucky Shop Tanjungbalai faces challenges in understanding diverse customer purchasing patterns, making it difficult to develop targeted marketing strategies. This study aims to apply the K-Means Clustering method to classify consumers based on purchasing behavior patterns. The data used consisted of 15 customer transaction records collected from Lucky Shop Tanjungbalai, with attributes including purchase frequency, quantity of purchased products, and product categories. This research adopted a qualitative approach combined with data mining techniques using the CRISP-DM framework, which consists of business understanding, data understanding, data preparation, modeling, evaluation, and deployment. The system was developed using PHP and MySQL. The results indicate that K-Means Clustering successfully segmented customers into Loyal Customers and Occasional Customers based on their purchasing characteristics. These segmentation results provide practical benefits for Lucky Shop by enabling more targeted promotional programs, improving customer relationship strategies, optimizing inventory planning, and supporting data-driven business decision-making. Therefore, the implementation of K-Means Clustering can serve as an effective solution for customer segmentation in local retail businesses.

Keywords:
Data Mining; K-Means Clustering; Consumer Segmentation; Marketing Strategy; PHP; MySQL.

1. INTRODUCTION

The rapid development of information technology has significantly influenced business processes in various sectors, especially in retail and trade industries. Companies are increasingly required to utilize technology to improve operational efficiency, enhance marketing strategies, and support decision-making processes. One of the technologies widely implemented for processing large-scale data is data mining. Data mining is a process used to discover hidden patterns, relationships, and valuable information from large datasets using specific analytical methods and algorithms (Gunawan & Purwayoga, 2022). Through data mining techniques, businesses can analyze customer behavior patterns and transform transaction data into strategic information that supports business growth and competitiveness.

In retail businesses, understanding customer purchasing behavior is an important factor in determining effective marketing strategies and inventory management. Customer purchasing patterns often vary depending on consumer preferences, shopping frequency, and product interests. Without proper analysis, companies may face problems such as ineffective promotional activities, inaccurate demand forecasting, and imbalance between inventory availability and market demand. Therefore, consumer segmentation becomes an important approach in identifying customer characteristics and grouping customers based on similarities in purchasing behavior. Consumer segmentation enables businesses to create targeted promotional strategies, improve customer satisfaction, and optimize stock management (Febriani & Putri, 2020).

One of the most widely used methods for consumer segmentation is K-Means Clustering. K-Means is a non-hierarchical clustering algorithm that groups data into several clusters based on similarity and distance calculations. This algorithm is considered effective because it can classify data automatically without predefined labels and produce clear customer groupings. The clustering process is generally performed by determining the number of clusters, calculating centroid values, measuring distances using Euclidean Distance, and iterating until stable cluster groups are formed (Gunawan & Purwayoga, 2022).

Several previous studies have demonstrated the effectiveness of K-Means Clustering in customer segmentation and marketing analysis. Research conducted by Perdana et al. (2022) showed that K-Means Clustering helps companies understand customer characteristics and improve customer loyalty through targeted marketing strategies in the Alfagift application. Another study by Wardani et al. (2023) comparing K-Means, DB Scanner, and Hierarchical Clustering methods concluded that K-Means provides stable clustering performance when data distribution is relatively balanced. Furthermore, Yunita et al. (2025) stated that customer segmentation using the RFM model combined with K-Means Clustering supports effective marketing decisions such as customer retention and reactivation strategies. Artiarno et al. (2025) also revealed that K-Means Clustering can identify purchasing patterns and support personalized marketing campaigns to increase customer loyalty in the retail sector.

Although many studies have implemented K-Means Clustering in retail and e-commerce sectors, most previous studies focused on large-scale companies and online marketplaces. Research related to clothing retail businesses at the regional distribution level, especially in local stores such as Lucky Shop Tanjungbalai, remains limited. This condition creates a research gap regarding the implementation of K-Means Clustering for analyzing customer purchasing patterns in small and medium-sized retail businesses. Therefore, this study offers both practical and academic contributions by implementing K-Means Clustering in consumer segmentation at Lucky Shop Tanjungbalai using historical customer transaction data.

Lucky Shop Tanjungbalai is a retail clothing store located on H.M. Nur Street No. 51, Tanjungbalai. The store experiences several challenges related to customer purchasing behavior because some customers purchase products regularly while others shop inconsistently. These conditions make it difficult for the store to determine effective marketing strategies and manage inventory efficiently. In addition, promotional activities and customer analysis are still conducted conventionally, resulting in less optimal business decisions. By implementing K-Means Clustering, customer transaction data can be grouped into several segments based on purchasing frequency, quantity of purchased products, and product categories. The clustering results are expected to help the store provide targeted promotions, improve customer service quality, and optimize inventory planning.

The novelty of this research lies in the implementation of K-Means Clustering for customer segmentation in a local clothing retail business environment using customer purchasing pattern attributes. Unlike previous studies that mainly focused on large-scale retail or e-commerce sectors, this study emphasizes the application of clustering techniques in small and medium-sized enterprises (SMEs) to support marketing and inventory management decisions. In addition, this research develops a web-based application using PHP and MySQL to facilitate the clustering process and customer data analysis (Putra & Muflih, 2024).

Based on the problems and research gaps identified above, this study aims to implement the K-Means Clustering method to classify consumers based on purchasing patterns at Lucky Shop Tanjungbalai. The results of this study are expected to improve marketing effectiveness, optimize inventory management, and support data-driven decision making in retail business operations.

2. RESEARCH METHOD

This study applies a qualitative research approach combined with data mining techniques to analyze customer purchasing patterns at Lucky Shop Tanjungbalai. The qualitative approach was selected because the research focuses on understanding business conditions, customer purchasing behavior, and the implementation of clustering methods in supporting marketing strategies and inventory management. The study also applies the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework because it provides systematic stages for conducting data mining processes (Gunawan & Purwayoga, 2022).

The data used in this study were obtained from historical customer transaction records at Lucky Shop Tanjungbalai during the observation period. A total of 15 customer transaction records were selected as research samples using purposive sampling techniques. The variables used in the clustering process consisted of purchase frequency, quantity of purchased products, and product categories. Data collection techniques included observation, interviews with store management, and documentation of transaction reports.

The data analysis process was conducted using the K-Means Clustering algorithm. The analysis began with data cleaning and transformation, followed by determining the number of clusters (K), initializing centroid values, calculating distances using the Euclidean Distance formula, assigning data to the nearest cluster, and recalculating centroid values iteratively until stable clusters were obtained. The resulting clusters were then interpreted to identify customer characteristics and support marketing decision-making.

3. RESULTS AND DISCUSSION

3.1. Implementation of K-Means Clustering

The implementation of the K-Means Clustering method in this study was carried out using customer transaction data obtained from Lucky Shop Tanjungbalai. The clustering process aimed to classify customers based on purchasing behavior patterns using several attributes, namely purchase frequency, quantity of purchased products, and product type. The implementation process consisted of determining the number of clusters, initializing centroid values, calculating distances using Euclidean Distance, grouping data into clusters, and performing iterative calculations until stable cluster results were obtained.

The initial stage of the clustering process was determining the centroid values for each cluster. The calculation process used Euclidean Distance to measure the similarity between customer data and centroid values. The distance calculation results for centroid C1 are shown in Table 1, while the calculations for centroid C2 and C3 are shown in Table 2 and Table 3.

Table 1. Centroid Iteration Value Data Table for C1

1	2	3	JUMLAH	C1 (AKAR)
9,000	49,000	9,000	67,000	8,185
81,000	361,000	16,000	458,000	21,401
4,000	25,000	4,000	33,000	5,745
121,000	576,000	16,000	713,000	26,702
1,000	9,000	1,000	11,000	3,317
49,000	196,000	9,000	254,000	15,937
16,000	81,000	4,000	101,000	10,050
100,000	441,000	16,000	557,000	23,601
0,000	4,000	1,000	5,000	2,236
36,000	169,000	9,000	214,000	14,629
64,000	289,000	9,000	362,000	19,026
25,000	121,000	4,000	150,000	12,247
1,000	16,000	1,000	18,000	4,243
49,000	225,000	9,000	283,000	16,823
4,000	36,000	4,000	44,000	6,633

Table 2. Centroid Iteration Value Data Table for C2

1	2	3	JUMLAH	C2 (AKAR)
64,000	16,000	64,000	144,000	12,000
4,000	64,000	49,000	117,000	10,817
81,000	36,000	81,000	198,000	14,071
0,000	169,000	49,000	218,000	14,765
100,000	64,000	100,000	264,000	16,248
16,000	9,000	64,000	89,000	9,434
49,000	4,000	81,000	134,000	11,576
1,000	100,000	49,000	150,000	12,247
121,000	81,000	100,000	302,000	17,378
25,000	4,000	64,000	93,000	9,644
9,000	36,000	64,000	109,000	10,440
36,000	0,000	81,000	117,000	10,817
100,000	49,000	100,000	249,000	15,780
16,000	16,000	64,000	96,000	9,798
81,000	25,000	81,000	187,000	13,675

Table 3. Centroid Iteration Value Data Table for C3

1	2	3	JUMLAH	C3 (AKAR)
64,000	16,000	64,000	144,000	12,000
4,000	64,000	49,000	117,000	10,817
81,000	36,000	81,000	198,000	14,071
0,000	169,000	49,000	218,000	14,765
100,000	64,000	100,000	264,000	16,248
16,000	9,000	64,000	89,000	9,434
49,000	4,000	81,000	134,000	11,576
1,000	100,000	49,000	150,000	12,247
121,000	81,000	100,000	302,000	17,378
25,000	4,000	64,000	93,000	9,644
9,000	36,000	64,000	109,000	10,440
36,000	0,000	81,000	117,000	10,817
100,000	49,000	100,000	249,000	15,780
16,000	16,000	64,000	96,000	9,798
81,000	25,000	81,000	187,000	13,675

The segmentation results indicate that customers were divided into two main groups, namely Loyal Customer and Occasional Customer. Customers categorized as Loyal Customer showed higher purchase frequency and product quantity compared to Occasional Customer. Loyal customers tended to make purchases regularly and contributed significantly to store sales. Meanwhile, Occasional Customers purchased products less frequently and had lower transaction intensity.

3.2. Additional Discussion

The clustering results indicate that Loyal Customers represent the most valuable customer segment because they contribute more consistently to store revenue. Therefore, Lucky Shop can implement customer retention programs such as membership cards, loyalty rewards, discount vouchers, and personalized promotions to maintain customer engagement.

On the other hand, Occasional Customers require different marketing approaches. Promotional campaigns, seasonal discounts, and product recommendations can be utilized to encourage repeat purchases and increase transaction frequency. The segmentation results also provide valuable information for inventory planning, as products frequently purchased by Loyal Customers can be prioritized to prevent stock shortages.

These findings support previous studies that emphasized the effectiveness of K-Means Clustering in identifying customer behavior patterns and improving marketing performance. Therefore, customer segmentation can become a strategic tool for enhancing competitiveness and operational efficiency in small retail businesses.

Table 4. Consumer Segmentation Data Table at Lucky Shop Tanjungbalai

No	Pelanggan	Tingkat Loyalitas
1	Rina Marlina	Loyal Customer
2	Andi Saputra	Occasional Customer
3	Budi Santoso	Loyal Customer
4	Citra Lestari	Occasional Customer
5	Dedi Pratama	Loyal Customer
6	Eka Putri	Occasional Customer
7	Fajar Nugroho	Loyal Customer
8	Gita Sari	Occasional Customer
9	Hendra Wijaya	Loyal Customer
10	Indah Permata	Occasional Customer
11	Joko Susanto	Occasional Customer
12	Kiki Amelia	Occasional Customer
13	Lukman Hakim	Loyal Customer
14	Maya Sari	Occasional Customer
15	Nanda Putra	Loyal Customer

3.3. System Implementation

The customer segmentation system was developed as a web-based application using PHP programming language and MySQL database. The system was designed to facilitate transaction data processing, customer data management, and clustering analysis automatically. The system implementation process included interface design, database implementation, and clustering module integration

The system provides two user roles, namely Admin and Owner. The admin is responsible for managing customer data, product data, transaction data, and executing the K-Means clustering process. Meanwhile, the Owner can view reports and customer segmentation results to support business decision making.

The database implementation consists of several tables, including users, customers, products, transactions, transaction details, and clustering results tables. The clustering results table stores customer segmentation information generated from the K-Means calculation process.

3.4. System Architecture

The system architecture consists of three main components: the user interface layer, application layer, and database layer. The user interface allows administrators and owners to interact with the system through web browsers. The application layer, developed using PHP, processes customer data, transaction records, and K-Means clustering calculations. The database layer utilizes MySQL to store customer information, product data, transaction histories, and clustering results. The interaction among these components enables automatic customer segmentation and report generation.

3.5. System Testing

System testing was conducted using the Black Box Testing method to evaluate whether the application functions operated correctly according to system requirements. The testing process focused on login functionality, customer data input, product management, transaction processing, clustering calculations, and report generation.

The testing results showed that all system features operated successfully and produced outputs according to user requirements. The clustering process successfully grouped customers automatically based on purchasing patterns. The system was also capable of displaying segmentation reports accurately and efficiently.

4. CONCLUSION

Based on the results of the research that has been conducted, the implementation of the K-Means Clustering method at Lucky Shop Tanjungbalai was successful in grouping consumers based on purchasing behavior patterns. The clustering process used several attributes, including purchase frequency, quantity of purchased products, and product types, to produce customer segmentation data effectively.

The results showed that consumers could be classified into several groups, namely Loyal Customer and Occasional Customer. Loyal customers were identified as customers with higher transaction frequency and purchasing intensity, while Occasional Customers showed lower purchasing activity. The implementation of customer segmentation can assist store management in understanding customer characteristics, improving marketing strategies, and optimizing inventory management.

In addition, the developed web-based system using PHP and MySQL successfully facilitated customer data management, transaction processing, and automatic clustering analysis. The system testing results indicated that all application features operated properly according to system requirements.

This study also demonstrates that the K-Means Clustering method is effective in supporting data-driven decision making in local retail businesses. The novelty of this research lies in the implementation of customer segmentation in a local clothing retail store environment to improve marketing effectiveness and operational efficiency.

From a practical perspective, the segmentation results can assist Lucky Shop in developing differentiated marketing strategies for each customer group. Loyal Customers can be targeted through loyalty programs and exclusive promotions, while Occasional Customers can be encouraged through discount campaigns and personalized product recommendations. Future studies are recommended to use larger datasets, integrate additional customer behavior variables, and compare the performance of K-Means with other clustering algorithms such as DBSCAN, Hierarchical Clustering, and Fuzzy C-Means to improve segmentation accuracy.

REFERENCES

A'yunan, Y. A. D. K., Indahyanti, U., & Busono, S. (2023). Implementasi Data Mining dalam Klasifikasi Diagnosa Kanker Payudara menggunakan Algoritma Logistic Regression. Jurnal Tekinkom (Teknik Informasi dan Komputer), 6(2), 400-407. https://doi.org/10.37600/tekinkom.v6i2.948

Djaka Permana, M., Lia Hananto , A. ., Novalia, E. ., Huda, B. ., & Paryono, T. . (2023). Klasterisasi Data Jamaah Umrah pada Tanurmutmainah Tour Menggunakan Algoritma K-Means. Jurnal KomtekInfo, 10(1), 15–20. https://doi.org/10.35134/komtekinfo.v10i1.332

Febriani, A., & Putri, S. A. (2020). Segmentasi Konsumen Berdasarkan Model Recency, Frequency, Monetary dengan Metode K-Means. JIEMS (Journal of Industrial Engineering and Management Systems), 13(2). http://dx.doi.org/10.30813/jiems.v13i2.2274

Gunawan, H., & Purwayoga, V. (2022). Data Mining Menggunakan Algoritma K-Means Clustering Untuk Mengetahui Potensi Penyebaran Virus Corona di Kota Cirebon. Jurnal Sisfokom (Sistem Informasi Dan Komputer), 11(1), 1–8. https://doi.org/10.32736/sisfokom.v11i1.1316

Heti Aprilianti, Khothibul Umam, & Maya Rini Handayani. (2025). Comparative Study of SVM, KNN, and Naïve Bayes for Sentiment Analysis of Religious Application Reviews. Journal of Applied Informatics and Computing, 9(3), 920–927. https://doi.org/10.30871/jaic.v9i3.9482

Hidayat, I., Askar, A., & Zaitun, Z. (2022). Teknologi Menurut Pandangan Islam. Prosiding Kajian Islam dan Integrasi Ilmu di Era Society (KIIIES) 5.0, 1(1), 456-460.

Iskandar, I., Triyanto, W. A., Fithri, D. L., & Arifin, M. (2025). Sistem Informasi Manajemen Stok dan Produksi Pakaian Anak Berbasis Web pada UMKM Linda Collection Menggunakan Metode Safety Stock. Jurnal SITECH: Sistem Informasi Dan Teknologi, 8(1), 105–116. https://doi.org/10.24176/sitech.v8i1.15556

Lim, M., & Ridho, M. R. (2021). RANCANG BANGUN SISTEM INFORMASI POINT OF SALE DENGAN FRAMEWORK CODEIGNITER PADA CV POWERSHOP. Computer and Science Industrial Engineering (COMASIE), 4(2), 46–55. Retrieved from https://ejournal.upbatam.ac.id/index.php/comasiejournal/article/view/3173

Nasrullah, A. H. (2021). Implementasi algoritma Decision Tree untuk klasifikasi produk laris. Jurnal Ilmiah Ilmu Komputer Fakultas Ilmu Komputer Universitas Al Asyariah Mandar, 7(2), 45-51. https://doi.org/10.35329/jiik.v7i2.203

Ordila, R., Wahyuni, R., Irawan, Y., & Sari, M. Y. (2020). Penerapan Data Mining Untuk Pengelompokan Data Rekam Medis Pasien Berdasarkan Jenis Penyakit Dengan Algoritma Clustering (Studi Kasus: Poli Klinik Pt. Inecda). Jurnal Ilmu Komputer, 9(2), 148-153. https://doi.org/10.33060/JIK/2020/Vol9.Iss2.181

Pudyawardana, W. (2023). Perancangan Sistem Informasi Pemesanan Makanan Dan Minuman Berbasis Web Pada Restoran Lamongan Cahaya. ALMUISY: Journal of Al Muslim Information System, 2(1), 21-27.

Ramdhan, D., Dwilestari, G., Dana, R. D., & Ajiz, A. (2022). Clustering data persediaan barang dengan menggunakan metode K-Means. MEANS (Media Informasi Analisa dan Sistem), 1-9. doi: 10.54367/means.v7i1.1826.

Soleh, O., & Jonas, D. (2024). Penggunaan Algoritma K-Means untuk Segmentasi Data Pelanggan pada Sistem Pemasaran Berbasis Data Mining. Nusa: Journal of Science Studies, 1(2), 63-69. https://doi.org/10.59613/0e5cay32

Srirahayu, A., & Pribadie, L. S. (2023). Review Paper Data Mining Klasifikasi Data Mining. Jurnal Ilmiah Informatika Global, 14(1). https://doi.org/10.36982/jiig.v14i1.2981

Utami, N., & Rosmita, R. (2024). Pengaruh Point of Purchase Terhadap Keputusan Pembelian Produk Pakaian di PT. Matahari Departement Store Ska Pekanbaru. JIABIS: Jurnal Ilmu Administrasi Bisnis dan Sosial, 2(2), 23-35.