



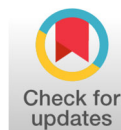
# Advanced Research Journal of Computer Science

Received: November 19, 2025 | Accepted: December 18, 2025 | Published: December 30, 2025  
Volume 02, Issue 02, Pages 09-13

DOI <https://doi.org/10.66590/arjcs2025020203>

Research Article

OPEN ACCESS



## An Intelligent Framework for Clustering-As-A-Service: A Case Study on Behavioral Segmentation

Rida Afzal<sup>1\*</sup>

<sup>1</sup>Department of Computer Science, University of Agriculture, Faisalabad, Pakistan

**Abstract** | The rapid growth of digital platforms and online services has significantly reduced the effectiveness of traditional socio-demographic segmentation methods, creating a need for more advanced and data-driven approaches. In response, this study proposes an intelligent support system designed to automate the evaluation of clustering processes within a Clustering-as-a-Service (CaaS) framework. The primary objective is to enhance customer behavior segmentation for improved decision-making, customer retention, and personalized service delivery. The proposed system is developed as part of a broader Customer Loyalty Intelligent Personalization (CLIP) platform, aiming to provide intelligent insights into customer patterns. A hierarchical clustering technique is employed to analyze real-world transactional data comprising 1,659 customers, 146 products, and 5,685 purchase records. Experimental results demonstrate that the proposed system effectively evaluates clustering performance and achieves high segmentation accuracy and efficiency. The findings highlight the potential of integrating intelligent evaluation mechanisms into clustering services to support scalable and data-driven customer analytics in modern business environments.

**Key Words** Customer Segmentation, Hierarchical Clustering, Clustering as a Service (CaaS)

Author Designation: <sup>1</sup>Researcher

\*Corresponding author: Rida Afzal (e-mail: [rida.99afzal@gmail.com](mailto:rida.99afzal@gmail.com)).

How to Cite the Article:

Afzal, Rida "An Intelligent Framework for Clustering-As-A-Service: A Case Study on Behavioral Segmentation." *Advanced Research Journal of Computer Science*, vol. 2, no. 2, December 2025, pp. 09-13. <https://doi.org/10.66590/arjcs2025020203>

### INTRODUCTION

In recent years, the relationship between organizations and customers has become a critical factor in achieving sustainable business success. Managing this relationship effectively requires structured strategies and intelligent systems, commonly referred to as Customer Relationship Management (CRM). CRM encompasses a range of techniques and tools designed to build long-term, profitable relationships by understanding customer needs, preferences, and behavior patterns.

With the rapid expansion of digital platforms and the increasing availability of online services, businesses are now dealing with a large and diverse customer base. This growth, combined with the rising complexity of customer behavior and the wide variety of available products, has made personalized recommendation and decision-support systems both essential and challenging. Traditional segmentation approaches based solely on socio-demographic factors are no longer sufficient to capture the dynamic nature of modern consumers.

Market segmentation remains a fundamental technique for organizing customers into meaningful groups, enabling targeted marketing strategies and improved customer engagement. These segments can be defined based on demographic, geographic, behavioral, or psychographic characteristics. However, the integration of data-driven approaches within CRM systems has significantly enhanced the effectiveness of segmentation. Intelligent CRM systems leverage data mining and machine learning techniques to extract valuable insights, optimize marketing campaigns, and improve customer retention rates.

Personalization has emerged as a key driver of customer satisfaction and loyalty. By analyzing customer interactions and purchase behavior, organizations can deliver tailored services and recommendations, ultimately strengthening their competitive advantage. In this context, the Customer Loyalty Intelligent Personalization (CLIP) system is introduced as an intelligent, machine learning-based framework that provides

real-time, adaptive recommendations for customer engagement. The system is designed to be flexible and applicable across various industries, allowing customization based on specific customer features and behavioral patterns.

This study focuses on the initial phase of the CLIP system, which involves automated customer segmentation based on behavioral and transactional data. Unlike traditional segmentation methods, the proposed approach considers both customer characteristics and purchasing patterns, enabling more accurate and meaningful grouping. To achieve this, an intelligent support system is developed to automate and evaluate the clustering process within a Clustering-as-a-Service (CaaS) environment.

The remainder of this paper is organized as follows: Section 2 reviews related work in customer segmentation and clustering techniques, Section 3 describes the proposed system architecture and methodology, Section 4 presents experimental results based on real-world customer data, and Section 5 concludes the study with key findings and future directions.

## Literature Review

Customer segmentation has been widely studied in the domain of data mining and machine learning, where researchers aim to identify meaningful patterns within large datasets to support decision-making and enhance customer engagement. The primary objective of these techniques is to group customers into distinct segments that enable organizations to better understand customer needs, optimize marketing strategies, and improve overall business performance.

Several studies have explored the application of Recency, Frequency, and Monetary (RFM) analysis combined with clustering algorithms for customer segmentation. For instance, You [1] developed a predictive model using RFM attributes along with K-means clustering and decision tree techniques to classify customers into meaningful groups. Their approach demonstrated the effectiveness of combining clustering with predictive analytics to support inventory management and marketing decisions.

Similarly, Abirami [2] proposed a customer classification framework integrating RFM analysis, K-means clustering, and association rule mining to analyze customer behavior in the retail sector. Their findings highlighted the importance of combining multiple data mining techniques to improve segmentation accuracy and uncover hidden relationships in customer data.

Comparative studies have also been conducted to evaluate the performance of different clustering algorithms. Shreya [3] analyzed K-means, hierarchical clustering, and hybrid approaches, concluding that K-means is more efficient for large datasets due to its computational simplicity, whereas hierarchical clustering is more suitable for smaller datasets with detailed structural analysis requirements.

In another study, Doğan [4] applied clustering techniques to segment a large customer dataset based on RFM values. Their results enabled businesses to categorize customers into premium and standard groups, allowing for targeted marketing strategies and improved customer retention. This study emphasized the practical importance of segmentation in real-world applications.

Advanced clustering approaches have also been explored to capture the dynamic nature of customer behavior. Yoseph [5] utilized both soft clustering (Fuzzy C-Means) and hard clustering (Expectation-Maximization) techniques to segment customers based on purchasing patterns. Their evaluation using cluster quality assessment metrics revealed that the EM algorithm performs better in terms of scalability and efficiency for certain datasets.

Further research by Yosepha [6] investigated the integration of clustering with regression models to analyze customer lifetime value and churn rates. Their findings demonstrated that data-driven segmentation strategies can significantly improve marketing outcomes, as evidenced by an increase in sales growth following the implementation of targeted campaigns.

Overall, existing literature highlights the critical role of clustering and data mining techniques in understanding customer behavior and enhancing marketing strategies. Most studies focus on identifying patterns, improving segmentation accuracy, and supporting business decision-making processes. However, limited attention has been given to automating the evaluation of clustering processes within a service-oriented framework.

To address this gap, the present study introduces an intelligent support system for automating clustering evaluation as part of a Clustering-as-a-Service (CaaS) environment. This work is integrated within a real-world e-commerce platform aimed at supporting Business-to-Consumer (B2C) applications. Customer segmentation serves as a foundational component of the proposed Customer Loyalty Intelligent Personalization (CLIP) system, where an intelligent mechanism is developed to enhance clustering performance and support personalized marketing strategies.

## Proposed Customer Segmentation System

This section presents the workflow and data processing pipeline of the proposed customer segmentation component. The system is designed as a structured, multi-stage process that transforms raw transactional data into meaningful customer segments through a sequence of four main phases.

The first phase involves Extract, Transform, and Load (ETL) operations, implemented as an independent web-based module. In this stage, transactional sales data is collected from users or external systems in CSV format. The ETL process converts and maps the raw data into a structured format compatible with the CLIP system,



Figure. 1: Customer Segmentation Expert System Dataflow

including the generation of a unified customer feature dataset. The second phase focuses on data preprocessing, where data validation and cleaning procedures are applied. This step ensures data quality by handling missing values, removing inconsistencies, and verifying data integrity before further analysis.

In the third phase, data aggregation and normalization are performed. Customer transactions are aggregated to form meaningful behavioral indicators, and normalization techniques are applied to standardize feature values. This process enhances the suitability of the dataset for clustering algorithms by ensuring balanced feature scaling.

The fourth and final phase involves customer segmentation using hierarchical clustering techniques. In this stage, each customer is represented as a feature vector, and clustering is performed to group customers based on similarity in their behavioral patterns. Hierarchical clustering is selected due to its ability to reveal data structure without requiring a predefined number of clusters (Figure 1).

The output of the system consists of well-defined customer segments, which are stored in a database for subsequent analysis and personalization processes within the CLIP framework.

The research uses real client data samples in the food sector. It consists of 1659 customers, 146 products, and 5685 orders. The development team selected and exported data that are related to sales only. The exported data consists of these main blocks, which were customer information, customer referrals, customer behavior, customer purchases, purchases per category, and loyalty offers. The research has aimed to automate the clustering phase to check if the data contains meaningful clusters or not.

The exported data has included customer features with minimum requirements, which were categorized as: behavioral and purchases features only. Initially, the development team cleaned these data and dumped the empty columns. Afterwards, hierarchical clustering is applied iteratively for defined sequential tasks in order to optimize and select the appropriate features to get customer clusters if they existed. Figure 2 shows the customer segmentation expert system flowchart. It works

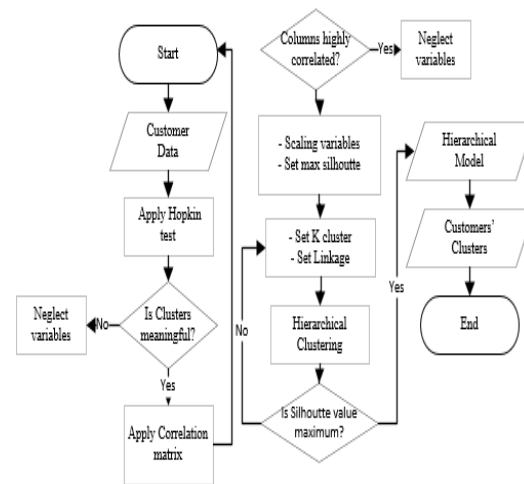


Figure 2: Customer Segmentation Flowchart

in two main steps: checking clusters' tendency and applying hierarchical clustering. Initially, Hopkins test [7] is applied to assure cluster tendency between feature columns, by testing the spatial randomness of the data. Next, correlation is measured to remove highly correlated columns. The preprocessed output will be the feature columns which are not empty nor highly correlated and non-uniformly distributed. Thereafter, the hierarchical clustering process works on clustering the preprocessed features. These features are scaled using minmax scalar [8] to normalize input features into range [0,1]. Consequently, the data scientists define the hierarchical clustering algorithm parameters, which are linkage [9] and k clusters. The linkage is used to measure similarity between clusters and k clusters is the number of defined clusters. The output segments/clusters are evaluated by silhouette score [10]. Repeatedly, those parameters and preprocessed features are varied and tuned to obtain the best combination to perform separable clusters. Briefly, the proposed minor expert system works as following:

- Apply Hopkin test on customer data to check features, if it forms meaningful clusters
- Apply correlation matrix on features to remove highly correlated features
- Scaling the selected feature columns
- Adjust hierarchical clustering algorithm parameters and train the selected features
- The build model is evaluated using silhouette score
- If the silhouette score is not maximum, then go to step 4
- The model selected if it achieves the maximum silhouette score

## RESULTS

### Experimental Results

There have been different applied experiments on real customer data. This section illustrates an experiment using client data in the food sector. It consisted of 1659 customers, 146 products, and 5685 orders. The team has

Table 1: Summary of Clustering Experiments (Rewritten)

Experiment	Category	Selected Features	Clustering Parameters	Evaluation (Silhouette Score)
Exp0	Customer Purchases	Visit Frequency, Total Orders, Morning Orders, Weekend Orders, Total Items, Total Purchase Amount, Avg. Order Value, Customer Lifespan, Order Frequency	Cluster 0 = 1658, Cluster 1 = 1, Linkage = Average	0.8532
	Purchases by Category	Category_1 to Category_5 Amount		
Exp1	Customer Purchases	Same as Exp0	Cluster 0 = 1657, Cluster 1 = 1, Cluster 2 = 1, Linkage = Average	0.7912
	Purchases by Category	Category_1 to Category_5 Amount		
Exp2	Customer Purchases	Same as Exp0	Cluster 0 = 319, Cluster 1 = 1340, Linkage = Ward	0.6887
	Purchases by Category	Category_1 to Category_5 Amount		
Exp3	Customer Purchases	Visit Frequency, Total Orders, Weekend Orders, Total Purchase Amount, Customer Lifespan, Avg. Order Value	Cluster 0 = 319, Cluster 1 = 1340, Linkage = Ward	0.6993

changed  $k$  clusters and linkage parameters to predict better clusters segments. The clustering has been evaluated using silhouette score. Table 1 represented the experiments' results. The table displayed in its four columns the experiment name, the main category of features, the detailed feature/column names, the defined parameters, and the evaluation metric respectively.

Table 1 presents a comparative summary of four clustering experiments conducted using different feature combinations and linkage methods. The evaluation of clustering quality is performed using the silhouette score, which measures the consistency and separation of clusters.

In Experiment 0, the clustering model achieved the highest silhouette score (0.8532), indicating strong cluster separation. However, the distribution of data points was highly imbalanced, with almost all customers grouped into a single cluster, limiting its practical usefulness.

Experiment 1 introduced an additional cluster, but the distribution remained highly skewed, with only a few data points assigned to minor clusters. Although the silhouette score remained relatively high (0.7912), the segmentation was not meaningful for real-world applications.

In Experiment 2, the use of Ward linkage produced more balanced clusters, with a significant distribution of customers across two groups. Although the silhouette score decreased (0.6887), the clustering results were more interpretable and practical for segmentation purposes.

Experiment 3 further refined the feature set by selecting the most relevant attributes. This resulted in a slight improvement in the silhouette score (0.6993) compared to Experiment 2, while maintaining balanced cluster distribution. This experiment demonstrates that feature selection plays a critical role in improving clustering performance.

Overall, the results indicate that while higher silhouette scores suggest better mathematical separation, balanced cluster distribution and interpretability are equally important for practical customer segmentation.

## CONCLUSIONS

This study is conducted as part of the development of a real-world e-commerce platform, namely the Customer

Loyalty Intelligent Personalization (CLIP) system. The primary objective of this research is to design and implement a portable and standalone Clustering-as-a-Service (CaaS) solution supported by an intelligent evaluation mechanism. The proposed system automates the selection and assessment of feature sets and hierarchical clustering parameters to generate meaningful and actionable customer segments without extensive manual intervention.

The developed framework is designed to be adaptable across different clients with minimal configuration requirements, thereby reducing the need for complex engineering and administrative efforts. By automating the clustering evaluation process, the system enhances the efficiency of the data science workflow and supports informed decision-making. The collaborative efforts of data scientists and researchers ensured that the system was rigorously tested using multiple real-world datasets to validate its reliability and effectiveness.

A case study conducted on real customer data from the food industry—comprising 1,659 customers, 146 products, and 5,685 transactions—demonstrated the practical applicability of the proposed approach. The system achieved a silhouette score of approximately 0.69, indicating a satisfactory balance between cluster cohesion and separation. Compared to traditional clustering methods, the proposed approach provides a more automated and scalable solution for customer segmentation. For future work, the system will be extended to support multiple application domains, with further improvements in feature engineering and data representation techniques. Additionally, the integration of generated customer segments into subsequent phases of the CLIP framework will enable enhanced personalization strategies aimed at increasing customer loyalty and reducing churn. The concept of automated model evaluation will also be expanded to include supervised learning techniques, allowing broader applicability in predictive analytics and intelligent decision support systems.

## REFERENCES

- [1] You, Z. *et al.* "A decision-making framework for precision marketing." *Expert Systems with Applications*, vol. 42, no. 7, 2015, pp. 3357–3367.
- [2] Abirami, M. *et al.* "Data mining approach for intelligent customer behavior analysis for a retail store." *Springer International Publishing*, vol. 49, 2016, pp. 283–291.
- [3] Tripathi, Shreya *et al.* "Approaches to clustering in customer segmentation." *International Journal of Engineering & Technology*, vol. 7, no. 3, 2018, pp. 802–807.
- [4] Doğan, O. *et al.* "Customer segmentation by using rfm model and clustering methods: A case study in retail industry." *International Journal of Contemporary Economics and Administrative Sciences*, vol. 8, no. 1, 2018, pp. 1–19.
- [5] Yoseph, F. *et al.* "New behavioral segmentation methods to understand consumers in retail industry." *International Journal of Computer Science & Information Technology (IJCSIT)*, vol. 11, no. 1, 2019.
- [6] Yosepha, F. *et al.* "The impact of big data market segmentation using data mining and clustering techniques." *Journal of Intelligent and Fuzzy Systems*, vol. 38, no. 5, 2020, pp. 6159–6173.
- [7] Banerjee, A. *et al.* "Validating clusters using the hopkins statistic." *IEEE International Conference on Fuzzy Systems*, 2004, pp. 149–153.
- [8] Patro, S. *et al.* "Normalization: A preprocessing stage." *International Advanced Research Journal in Science, Engineering, and Technology*, vol. 2, no. 3, 2015.
- [9] Yim, O. *et al.* "Hierarchical cluster analysis: Comparison of three linkage measures and application to psychological data." *The Quantitative Methods for Psychology*, vol. 11, no. 1, 2015, pp. 8–21.
- [10] Rousseeuw, P.J. "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis." *Journal of Computational and Applied Mathematics*, vol. 20, 1987, pp. 53–65.