TELKOMNIKA Telecommunication Computing Electronics and Control
Vol. 23, No. 1, February 2025, pp. 119~128
ISSN: 1693-6930, DOI: 10.12928/TELKOMNIKA.v23i1.26384

Customer segmentation in e-commerce: K-means vs hierarchical clustering

Sumit Kumar[1], Ruchi Rani[2], Sanjeev Kumar Pippal[3], Riya Agrawal[2]
[1]Symbiosis Institute of Technology, Pune Campus, Symbiosis International (Deemed University), Pune, India
[2]Department of Computer Engineering and Technology, School of Computer Engineering and Technology, Dr. Vishwanath Karad MIT World Peace University, Pune, India
[3]GL Bajaj Institute of Technology and Management, Greater Noida, India

Article Info

Article history:

Received Jun 5, 2024
Revised Aug 11, 2024
Accepted Sep 20, 2024

Keywords:

Brand interests
Customer insights
Customer segmentation
E-commerce
Hierarchical clustering
K-means clustering

ABSTRACT

Customer segmentation is important for e-commerce companies to understand and target different customers. The primary focus of this work is the application and comparison of K-means clustering and hierarchical clustering, unsupervised machine learning techniques, in customer segmentation for e-commerce platforms. Clustering leverages customer search behavior, reflecting brand preferences, and identifying distinct customer segments. The proposed work explores the K-means algorithm and hierarchical clustering. It uses them to classify customers in a standard e-commerce customer dataset, mainly focused on frequently searched brands. Both techniques are compared based on silhouette scores and cluster visualizations. K-means clustering yielded well-separated segments compared to hierarchical clustering. Then, using the K-means algorithm, customers are classified into different segments based on brand search patterns. Further, targeted marketing strategies are discussed for each segment. Results show three customer segments: high searchers-low buyers, loyal customers, and moderate engagers. The proposed work provides valuable insights into customers that could be used for developing targeted marketing campaigns, product recommendations, and customer engagement strategies to enhance the conversion rate, customer satisfaction, and, in turn, the growth of an e-commerce platform.

Corresponding Author:

Ruchi Rani
Department of Computer Engineering and Technology, School of Computer Engineering and Technology
Dr. Vishwanath Karad MIT World Peace University
Pune, Maharashtra, India
Email: ruchiasija20@gmail.com

1. INTRODUCTION

In the highly competitive e-commerce world, where customer preferences and behaviors continually evolve, understanding the customer base has become a necessity rather than merely a choice [1]. There is no room for the traditional generic marketing methods today. This requires a more granular approach, where customer segmentation assumes paramount importance. Based on shared traits like demographics, past purchases, or browsing habits, customer segmentation creates groupings of customers. This allows e-commerce companies to fine-tune their marketing strategies and product recommendations to best appeal to each segment. The study will dive into unsupervised machine learning techniques for customer segmentation in e-commerce: K-means clustering and hierarchical clustering. K-means clustering groups data points (customers in this case) together on a similarity score based on features (search behavior in this case) [2]. Emphasis is placed on analyzing customer search behavior, which strongly indicates specific brand preferences and purchase intent.

The analysis of search patterns will lead to the creation of distinct customer segments with distinctive characteristics. The study employs K-means clustering and compares it with another common technique, hierarchical clustering. K-means clustering outperforms hierarchical clustering in identifying well-separated customer segments based on e-commerce search behavior. K-means clustering produced a higher silhouette score (0.29268) than hierarchical clustering (0.25372), indicating better-defined and more distinct segments. Next, K-means is implemented on a real e-commerce customer dataset focusing on the most frequently searched brands. The data is preprocessed, and then customers are grouped to uncover distinct segments. The study discusses targeted marketing strategies for each segment to boost conversion rates, customer satisfaction, and the overall growth of an e-commerce platform.

The structure of this manuscript is as follows: section 2 gives an encompassing literature review of the existing works on customer segmentation using machine learning techniques and the main findings from related research efforts. Section 3 presents the methodology with a description of the dataset, preprocessing steps for data, and the clustering algorithms applied for customer segmentation. The K-means and hierarchical clustering techniques have been compared based on silhouette scores and visualizations in section 4. Finally in section 5, conclusions are drawn about the research work and suggestions for future work that could enrich customer segmentation methods in e-commerce platforms.

2. LITERATURE REVIEW

Customer segmentation is considered the heart of any e-commerce business, wherein it can apply its strategies to more diversified needs of different customers. Various research on different customer segmentation techniques for e-commerce platforms have been performed in this regard. Rajput and Singh [2] used K-means clustering with purchase data as attributes for customer segmentation. It was efficient enough to find a specific pattern of purchase behavior. Similarly, Tabianan *et al*. [3] leveraged K-means for customer segmentation using purchase behavior data and highlighted that the techniques worked well to identify customer groups with unique buying behavior. Beyond the analysis of purchase data, other customer segmentation factors include: i) recency, frequency, and monetary (RFM) analysis: as outlined by Kaymak [4], RFM analysis segments customers according to the RFM value of their purchases. However, the RFM model might lack a forward-looking perspective for customer value assessment. To address this, integrating machine learning and predictive analytics could enhance RFM models, enabling more accurate predictions, and personalized marketing strategies in e-commerce; ii) demographic segmentation: this segmentation factor groups customers based on age, gender, income, and location; and iii) psychographic segmentation: it divides the customer base according to lifestyle, interests, values, and personality traits.

Though these techniques provide valuable insights, K-means clustering provides a data-driven approach that directly uses the customer’s behavior as input. Our study focuses on customer search behavior as a segmentation factor. We will emphasize analyzing the search pattern, which will help reveal customers with similar brand preferences and purchase intent. This will help develop more targeted marketing strategies than traditional segmentation methods. Prior research on e-commerce customer segmentation has focused chiefly on purchase history data, as shown in Table 1 [5]-[8].

Table 1. Comparison of literature work

Ref.	Focus	Dataset/attributes	Methods/algorithms	Key contribution
[5]	Customer segmentation using clustering algorithms to guide effective marketing and product development.	Multiple datasets with varying attributes.	Clustering algorithms, latent class model.	Visualization of customer segments and latent class model to handle mixed-type measures.
[6]	Telecom customer segmentation using the FTCA model and improved density peak clustering algorithm for better classification.	Telecom customer data, including number of calls, average call time, billing, and accumulated arrears.	Frequency, time, cost and amount (FTCA) model, adaptive density peak clustering with Gini coefficient.	Introduces an adaptive density peak algorithm to improve telecom customer segmentation and classification accuracy.
[7]	E-commerce customer segmentation uses K-means to analyze customer-related parameters like date, customer ID, product category, and payment method.	E-commerce site data include date, customer ID, product category, payment method, value, time on site, and clicks.	K-means clustering.	Explores customer segmentation basics and their importance in the digital market, focusing on personalized experiences.
[8]	Addressing information overload in e-commerce by using customer segmentation and personalization techniques.	Clickstream data, transaction data, demographic and psychographic attributes.	K-means clustering and support vector regression (SVR)	Combines customer segmentation with SVR to enhance personalization and customer targeting in e-commerce.

Though informative, purchase data reflects past behavior and might not be good enough to predict future actions. Our study explores customer search behavior as a segmentation factor to fill this gap. Search data provides customer intent and brand preference information, which might predict future purchase decisions. We are trying to find customer segments with a higher propensity to purchase particular brands or product categories by analyzing the search patterns. This way, we focus on search behavior for more proactive and targeted marketing strategies, which might raise the conversion rate and customer satisfaction.

3. METHOD

This study explores customer search behavior as a segmentation factor. Emphasis on analyzing the search pattern will help reveal customers with similar brand preferences and purchase intent, which might predict future purchase decisions. This approach will allow for more precise targeted marketing strategies than traditional segmentation methods, which often struggle to handle the complexities of big data, thus limiting their effectiveness [9]. This study aims to find customer segments with a higher propensity to purchase particular brands or product categories by analyzing the search patterns. It allows for more proactive and targeted marketing, which might raise the conversion rate and customer satisfaction.

3.1. Dataset description

The dataset comprises 30,000 entries with 38 columns representing customer behavior and demographic attributes. Cust_ID: this column serves as the primary key for identification. This is an integer unique to every customer. Gender: it's a column of categorical data representing the customer's gender. However, this column contains some missing entries. Orders: integer variable containing total orders by the respective customer. The remaining 35 columns in this data set are integer columns, which denote the count of purchased products by the customer from different brands like Jordan, Samsung, H&M, and Microsoft. These counts usually range from zero to up to a few purchases and reflect customers' shopping preferences. The data here can present a complete picture of consumer behavior across multiple brands and is useful for purchase pattern analysis or segmentation of customers.

3.2. Data preprocessing

Before applying clustering algorithms, data preprocessing is crucial to ensure data quality and prepare it for analysis. The steps include:

- Handling missing values: missing values can be handled by imputing the mode, which is the most frequent value. Alternative approaches for missing value handling exist, such as removing rows with missing entries or applying imputation techniques based on data distribution and the amount of missingness, depending on the specific characteristics of the data [10].
- Scaling of data: scaling has its place in ensuring that each feature equally contributes to the clustering process. This stops larger-scale features from influencing the process [11]. MinMaxScaler, available from the scikit-learn library, can scale the brand search frequency features between 0 and 1.

3.3. K-means clustering for customer segmentation

K-means is an unsupervised machine-learning method that looks for patterns and clusters within a data set. The algorithm randomly selects k centroids, the centers of the clusters, assigns the data points to the nearest centroids, and updates them to an average of the data points in the cluster. This process repeats until the centroids stop changing [11].

The determination of the best K value: it is highly critical in the process of obtaining effective customer segmentation. The study used the following two methods to obtain the optimal number of clusters:

- The elbow method: the elbow method explores the trade-off between the number of clusters and the explained variance of each cluster [12]. K-means clustering is executed for various K values, and for each cluster configuration, the inertia is calculated. As shown in Figure 1, an elbow curve is formed, and the "elbow" point on the curve represents the ideal number of clusters, K. From that point forward, adding more clusters will not explain much more variance to avoid overfitting the data.
- The silhouette analysis: silhouette analysis studies the quality of clustering based on how well each data point is assigned to its assigned cluster compared to how close it is to points in other clusters. As shown in Figure 2, the silhouette scores range from -1 to 1, where a higher value closer to 1 indicates better separation between clusters and proper assignment of data points to their respective clusters. The average silhouette score on all data points is calculated for each K value.

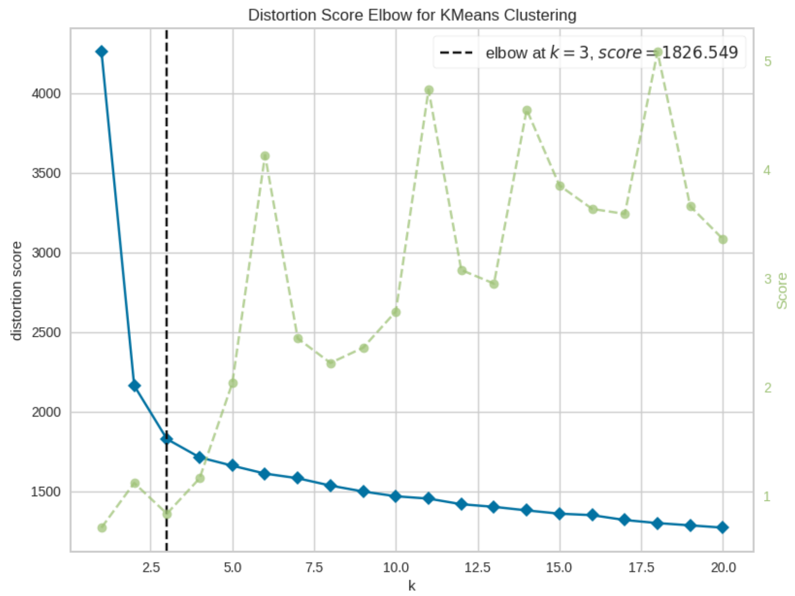

Figure 1. Elbow point curve for the optimal number of k

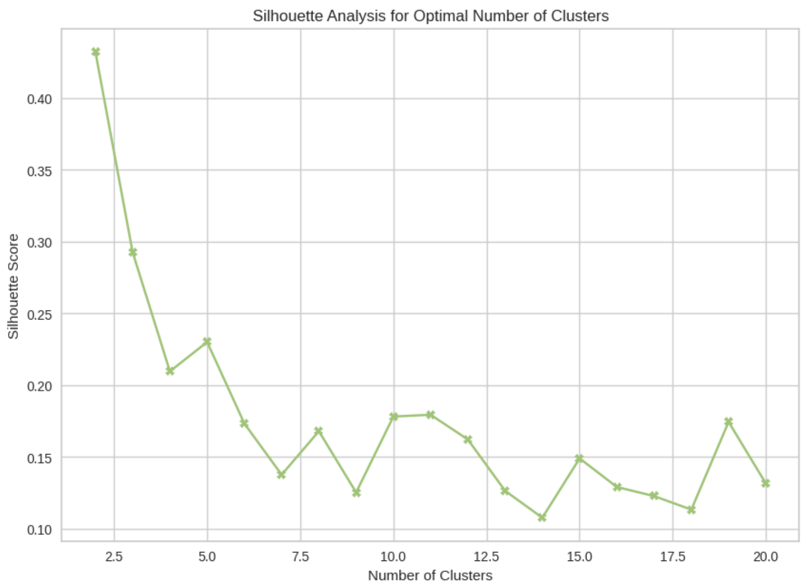

Figure 2. Silhouette curve for optimal no. of k

K-means clustering allows e-commerce users to be classified into subsets based on the customers' search history information. The process involves several key steps:

- Data preparation: the customer data is preprocessed to facilitate cluster analysis, involving the following steps:
  a. Handling missing values: the missing values are managed by adding default values or removing rows with incomplete data [2].
  b. Ensuring feature equality: to ensure that every feature has the same weight in the cluster analysis, numerical features, like the number of previous orders, are scaled. Additionally, categorical data such as gender is encoded using techniques like one-hot encoding.
- Feature selection: the process involves the removal of unnecessary features [13], with emphasis placed on selecting features relevant to customer segmentation aligned with business goals. For instance, "Cust_ID" and "Orders" may be excluded if they do not reflect brand search behavior directly. Instead, priority is given to columns indicating brand search frequency.

- Determining the optimal number of clusters (K): this is a very important step that significantly impacts the segmentation results [14]. Methods like domain knowledge, the elbow method, and silhouette analysis are considered to determine the ideal number of clusters based on customer groups such as brand devotees or price-conscious shoppers.
- Apply K-means clustering: the K-means clustering technique is applied to the customer data based on the selected K value.
- Interpret the clusters: interpret the resultant customer groups through the following steps:
- Characterize cluster properties: analyzing the average brand search frequency within each cluster provides valuable insights into the primary brand preferences of each customer group. This step is crucial for understanding the distinct characteristics of each cluster in the segmentation analysis.
- Cluster visualization: depending on the number of clusters formed and the data characteristics, utilize dimensionality reduction techniques such as scatterplots or principal component analysis (PCA) to visualize the data in a lower-dimensional space [2], [15].

The K-means clustering algorithm is performed based on insights from the elbow method and silhouette analysis, classifying each customer into one of the K clusters. Thus, customer data grouping is based on their search behavioral pattern.

3.4. Hierarchical clustering

It is another unsupervised machine-learning algorithm that can be used for customer segmentation. The dendrogram, a tree-like diagram highlighting the relationship between each data point, is built through this algorithm [16]. Natural clusters within the data can then be identified from the resulting dendrogram. This is the way it operates:

- Data preprocessing: prepare and scale numerical features in customer data to ensure they are suitable for distance calculations.
- Choosing a linkage method: select an appropriate linkage method according to your data characteristics and desired clustering outcomes [17].
- Execute agglomerative clustering: gradually combine similar customer clusters using the chosen link technique.
- Analysis of dendrogram: the generated dendrogram visually displays clusters' merging steps and hierarchical links. Agglomerative refers to a "bottom-up" approach that combines individual data points into clusters [18].
- Define customer segments: examine the dendrogram to pinpoint significant customer segments according to the preferred level of detail.

Agglomerative hierarchical clustering is performed on the preprocessed data to compare the results with K-means clustering. This method, also known as Ward's minimum variance method, utilizes Ward's linkage and the Euclidean distance metric to build a hierarchical cluster structure by minimizing the total variance within each cluster at each stage, resulting in high intra-cluster similarity (objects within a cluster are very similar) [19]. Two different ways of analysis are used to assess the quality of customer segmentation: silhouette score and cluster visualization [20]. Silhouette scores for K-means and hierarchical clustering are calculated to determine if both methods generated well-separated clusters and to evaluate which clustering technique is more suitable.

4. RESULTS AND DISCUSSION

Comparison of silhouette scores: the silhouette scores for K-means and hierarchical clustering are compared to examine the quality of the cluster separation. A higher average silhouette score indicates better-defined and distinct segments [21]. As shown in Figure 3, the silhouette score for K-means was 0.29268, whereas the hierarchical clustering achieved a score of 0.25372. Since higher values of silhouette score close to 1 indicate better separation, K-means clustering produced a slightly higher score, suggesting that K-means produced more distinctive clusters.

```
Silhouette Score for K-means Clustering: 0.2926824214520362
Silhouette Score for Hierarchical Clustering: 0.2537253019218621
```

Figure 3. Comparison of silhouette score for K-means clustering and hierarchical clustering

Cluster visualization: visualization of cluster assignments for both methods, as shown in Figure 4, supports that K-means produced more separated clusters in the feature space than hierarchical clustering. This visual separation indicates that K-means may have yielded better-defined customer segments based on the search behavior patterns. Additionally, the tighter and more consistent grouping of points in K-means clusters reinforces the notion of clearer distinctions among segments.

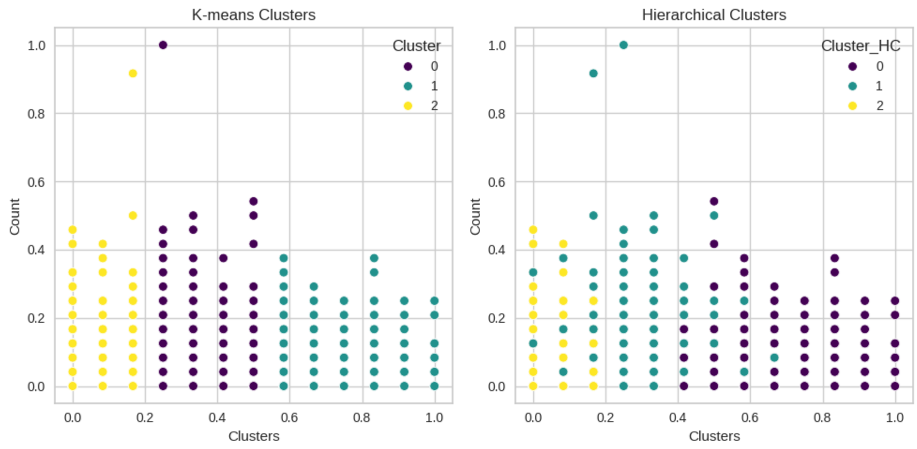

Figure 4. Comparison of clusters formed by both methods

Visualization of K-means clusters: the clustering of customers using K-means is illustrated in Figure 5. If visualizing all features of brand search is not feasible, the first two principal components can be used for dimensionality reduction [2], [22]. This makes the distribution pattern among customers in K-means clusters visually apparent.

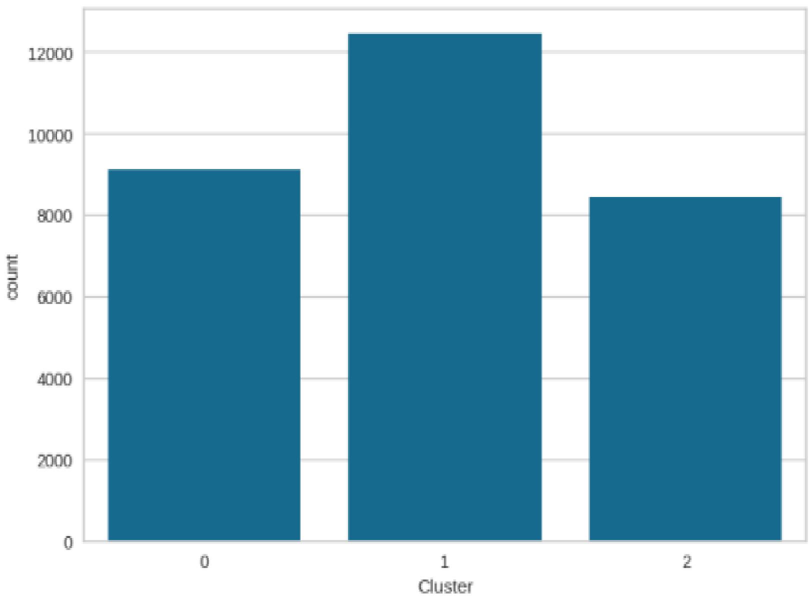

Figure 5. Assignment of data points to clusters for K-means clustering

Customer segment analysis: the characteristics of each segment are analyzed to understand their search behavior and potential purchase preferences. Some key metrics are analyzed, like average order number, the total number of times a customer performed a search, and gender distribution within a segment to define

different profiles of customers. Through data preprocessing, K-means clustering, evaluation techniques, and visualization, e-commerce customer segments can be identified based on search behavior [23]. It will help define various customer profiles, unique preferences, and the purchase propensity of brands, facilitating targeted marketing strategies. Based on the results obtained for the K-means clustering as shown in Figure 6, three different customer segments were identified, each with unique characteristics:
- Cluster 0 (high searchers, high buyers): customers with moderate search activity (43,966 searches) but a high order volume (79,885 orders). This indicates they research before purchasing, suggesting a high-value customer segment.
- Cluster 1 (high searchers, low buyers): this cluster is characterized by the highest search activity (56,009) with a relatively low order volume (7,560). These customers likely research extensively before purchasing but convert at a lower rate than other clusters. Understanding why they search so much but buy less can be crucial for targeted marketing efforts.
- Cluster 2 (moderate engagement): customers in this cluster exhibit moderate search activity (40,795) and order volume (37,649). Their behavior falls between the high-engagement and loyal customer segments.

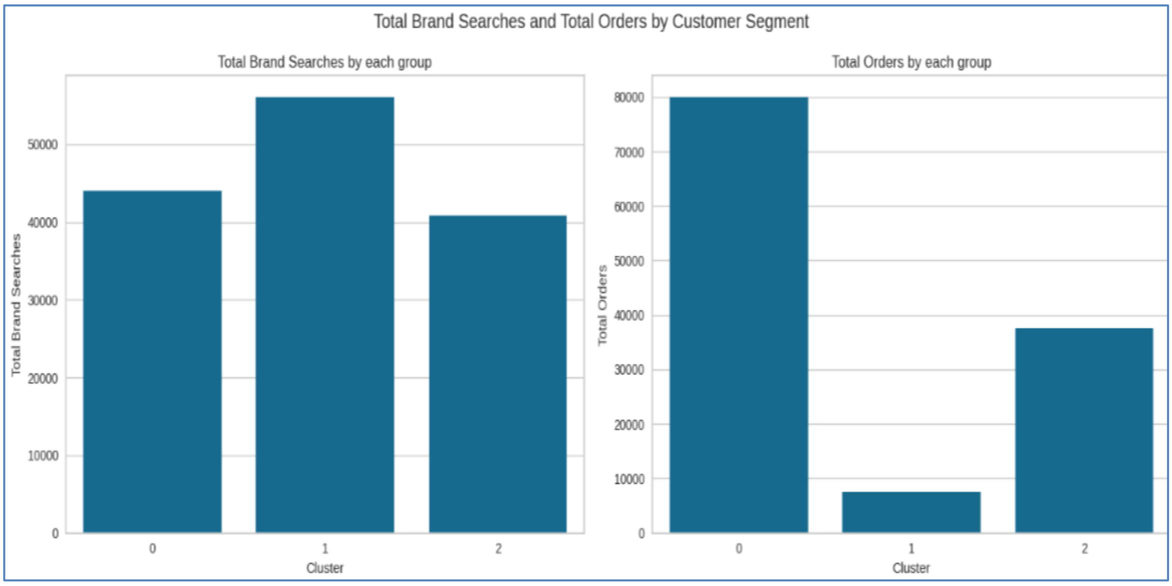

Figure 6. Cluster analysis for K-means clustering

It is worth noting that while k-means clustering is a popular method for customer segmentation, it can encounter challenges in high-dimensional settings with correlated variables [24]. When multiple correlated features are present, the distance metrics used for clustering can get distorted, making it difficult for the algorithm to identify separate clusters. Therefore, preprocessing steps like dimensionality reduction or feature selection are important to improve the effectiveness of K-means clustering.

Marketing strategies for customer segments: this customer segmentation provides marketing professionals with a powerful tool. It allows them to develop targeted messaging and campaigns that resonate with specific customer groups, leading to short-term wins and long-term customer loyalty [25]. By understanding the distinct characteristics of each customer segment, such targeted marketing strategies can be developed:
- High searchers, low buyers: implement targeted promotions, special offers, or personalized recommendations based on their search history to convert browsing activity into purchases.
- Loyal customers: employ loyalty programs, exclusive discounts, or personalized recommendations for complementary products to retain these valuable customers and potentially encourage higher-order values.
- Moderate engagement: use A/B testing of product offerings and targeted advertising campaigns informed by their search behavior and past purchases to personalize their shopping experience and convert them into more frequent buyers.

Businesses can improve customer engagement, conversion rates, and overall customer lifetime value by tailoring marketing efforts to each segment's needs and behavior. Table 2 compares our work with the other work done in the literature.

Table 2. Comparison of our work with literature work

Ref.	Focus	Dataset/attributes	Methods/algorithms	Key contribution
[26]	Latent semantic user segmentation for online advertising using user search behaviors to improve ad targeting.	Search behaviors include queries issued and ads clicked by users.	Latent dirichlet allocation (LDA).	Proposes a latent semantic segmentation strategy that achieves higher click-through rate (CTR) by grouping users based on search behaviors.
[27]	Multichannel consumer behavior analysis across store, internet, mobile, and social media channels using latent-class cluster analysis.	Consumer behavior data across different retail channels, including clothing, holiday travel, and consumer electronics.	Latent-class cluster analysis.	Identifies five multichannel consumer segments, showing the evolution of consumer preferences and behavior in retail channels.
[28]	Enhancing purchase prediction and customer segmentation in e-commerce by integrating customer characteristics with browsing patterns.	Clickstream data, transaction data, RFM approach, graph metrics for browsing patterns.	RFM analysis, graph metrics, predictive analysis, clustering analysis.	Proposes a novel session-level customer segmentation approach that significantly improves purchase prediction by integrating customer characteristics and browsing patterns.
Our work	Application and comparison of K-means and hierarchical clustering in e-commerce customer segmentation, focusing on search behavior and brand preferences.	Standard e-commerce customer dataset, focusing on frequently searched brands.	K-means clustering, hierarchical clustering, silhouette scores, cluster visualizations.	Demonstrates that K-means clustering yields better-separated segments, identifies three customer segments (high searchers-low buyers, loyal customers, moderate engagers), and provides actionable insights for targeted marketing strategies, product recommendations, and customer engagement.

5. CONCLUSION

This research employed customer segmentation analysis to identify distinct customer groups based on their e-commerce search behavior. K-means clustering emerged as the preferred approach, revealing three distinct customer segments: (high searchers, low buyers), loyal customers, and moderate engagement. Each segment exhibits unique characteristics, presenting valuable opportunities for targeted marketing campaigns, product recommendations, and improved customer engagement strategies. Future research could explore the inclusion of additional customer data and investigate the effectiveness of alternative clustering techniques to refine customer segmentation strategies further. Future work could benefit from employing a broader array of clustering techniques beyond those used in this study. Techniques such as density-based spatial clustering of applications with noise (DBSCAN), Gaussian mixture models (GMM), and self-organizing maps (SOM) could provide additional insights and improved segmentation. Additionally, further analysis could focus on differentiation of the specific product categories or customer demographics within each segment. This approach would help gain more nuanced insights into customer behavior for better-targeting marketing campaigns, product recommendations, and other customer engagement initiatives. Such efforts could enhance conversion rates and customer satisfaction, driving overall business growth for e-commerce platforms.

REFERENCES

[1] N. Saini, K. Sharma, P. K. Sarangi, G. Singh, and L. Rani, "Customer Segmentation using K-Means Clustering," *2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, Noida, India, 2022, pp. 1-5, doi: 10.1109/ICRITO56286.2022.9965147.

[2] L. Rajput and S. N. Singh, "Customer Segmentation of E-commerce data using K-means Clustering Algorithm," *2023 13th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Noida, India, 2023, pp. 658-664, doi: 10.1109/Confluence56041.2023.10048834.

[3] K. Tabianan, S. Velu, and V. Ravi, "K-Means Clustering Approach for Intelligent Customer Segmentation Using Customer Purchase Behavior Data," *Sustainability*, vol. 14, no. 12, 2022, doi: 10.3390/su14127243.

[4] U. Kaymak, "Fuzzy target selection using RFM variables," *Proceedings Joint 9th IFSA World Congress and 20th NAFIPS International Conference (Cat. No. 01TH8569)*, Vancouver, BC, Canada, vol. 2, pp. 1038-1043, 2001, doi: 10.1109/NAFIPS.2001.944748.

[5] S. Kaur and Sarabjeet, "Customer Segmentation Using Clustering Algorithm," *2021 International Conference on Technological Advancements and Innovations (ICTAI)*, Tashkent, Uzbekistan, 2021, pp. 224-227, doi: 10.1109/ICTAI53825.2021.9673169.

[6] X. Pu, C. Song, and J. Huang, "Research on Optimization of Customer Value Segmentation Based on Improved K-Means Clustering Algorithm," *2020 IEEE 3rd International Conference on Information Systems and Computer Aided Education (ICISCAE)*, Dalian, China, 2020, pp. 538-542, doi: 10.1109/ICISCAE51034.2020.9236867.

[7] R. Punhani, V. P. S. Arora, S. Sabitha, and V. K. Shukla, “Application of Clustering Algorithm for Effective Customer Segmentation in E-Commerce,” *2021 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*, Dubai, United Arab Emirates, 2021, pp. 149-154, doi: 10.1109/ICCIKE51210.2021.9410713.

[8] C. S. Vamsee, D. Rakesh, I. Prathyusha, B. Dinesh, and C. Bharathi, “Demographic and Psychographic Customer Segmentation for Ecommerce Applications,” *2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, Salem, India, 2023, pp. 615-622, doi: 10.1109/ICAAIC56838.2023.10140861.

[9] S. Priyadarshni, R. Fathima, S. Urolagin, A. M. Bongale, and D. S. Dharrao, “Unveiling Customer Segmentation Patterns in Credit Card Data using K-Means Clustering: A Machine Learning Approach,” *2023 International Conference on Modeling, Simulation & Intelligent Computing (MoSICom)*, Dubai, United Arab Emirates, 2023, pp. 362-366, doi: 10.1109/MoSICom59118.2023.10458783.

[10] S. Alam, M. S. Ayub, S. Arora, and M. A. Khan, “An investigation of the imputation techniques for missing values in ordinal data enhancing clustering and classification analysis validity,” *Decision Analytics Journal*, vol. 9, p. 100341, 2023, doi: 10.1016/j.dajour.2023.100341.

[11] S. U. Masruroh, A. T. Fadilah, K. Hulliyah, A. F. Ramadhan, R. A. Putri, and M. A. Irfan, “Implementation of the K-means Clustering Algorithm for Targeting Ads. Case Study: IBM Watson Analytics Car Insurance Customer Data,” *2023 11th International Conference on Cyber and IT Service Management (CITSM)*, Makassar, Indonesia, 2023, pp. 1-3, doi: 10.1109/CITSM60085.2023.10455449.

[12] B. Kaur and J. R. Saini, “A Strategy to Identify Loyalty Using Elbow Curve Method for Customer Segmentation,” *2022 IEEE Pune Section International Conference (PuneCon)*, Pune, India, 2022, pp. 1-7, doi: 10.1109/PuneCon55413.2022.10014742.

[13] S. Chormunge and S. Jena, “Correlation based feature selection with clustering for high dimensional data,” *Journal of Electrical Systems and Information Technology*, vol. 5, no. 3, pp. 542-549, 2018, doi: 10.1016/j.jesit.2017.06.004.

[14] P. Patel, B. Sivaiah, and R. Patel, “Approaches for finding Optimal Number of Clusters using K-Means and Agglomerative Hierarchical Clustering Techniques,” *2022 International Conference on Intelligent Controller and Computing for Smart Power (ICICCSP)*, Hyderabad, India, 2022, pp. 1-6, doi: 10.1109/ICICCSP53532.2022.9862439.

[15] U. Sharma, G. Aditi, N. R. Roy, and S. N. Singh, “Analysis of Customer Segmentation Clustering Techniques,” *2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Noida, India, 2022, pp. 374-379, doi: 10.1109/Confluence52989.2022.9734147.

[16] M. Roux, “A Comparative Study of Divisive and Agglomerative Hierarchical Clustering Algorithms,” *Journal of Classification* vol. 35, pp. 345–366, 2018, doi: 10.1007/s00357-018-9259-9.

[17] Vijaya, S. Sharma, and N. Batra, “Comparative Study of Single Linkage, Complete Linkage, and Ward Method of Agglomerative Clustering,” *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, Faridabad, India, 2019, pp. 568-573, doi: 10.1109/COMITCon.2019.8862232.

[18] C. Gopal and L. Jacob, “Customer Behavior Analysis Using Unsupervised Clustering and Profiling: A Machine Learning Approach,” *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, Greater Noida, India, 2022, pp. 2075-2078, doi: 10.1109/ICACITE53722.2022.9823646.

[19] P. P. Pramono, I. Surjandari, and E. Laoh, “Estimating Customer Segmentation based on Customer Lifetime Value Using Two-Stage Clustering Method,” *2019 16th International Conference on Service Systems and Service Management (ICSSSM)*, Shenzhen, China, 2019, pp. 1-5, doi: 10.1109/ICSSSM.2019.8887704.

[20] N. Tomašev and M. Radovanović, “Clustering Evaluation in High-Dimensional Data,” *Unsupervised Learning Algorithms. Springer, Cham*, pp. 71-107, 2016, doi: 10.1007/978-3-319-24211-8_4.

[21] K. R. Shahapure and C. Nicholas, “Cluster Quality Analysis Using Silhouette Score,” *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, Sydney, NSW, Australia, 2020, pp. 747-748, doi: 10.1109/DSAA49011.2020.00096.

[22] K. Z. Mao, “Identifying critical variables of principal components for unsupervised feature selection,” in *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 35, no. 2, pp. 339-344, Apr. 2005, doi: 10.1109/TSMCB.2004.843269.

[23] T. Kansal, S. Bahuguna, V. Singh, and T. Choudhury, “Customer Segmentation using K-means Clustering,” *2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)*, Belgaum, India, 2018, pp. 135-139, doi: 10.1109/CTEMS.2018.8769171

[24] H. -H. Zhao, X. -C. Luo, R. Ma, and X. Lu, “An Extended Regularized K-Means Clustering Approach for High-Dimensional Customer Segmentation with Correlated Variables,” in *IEEE Access*, vol. 9, pp. 48405-48412, 2021, doi: 10.1109/ACCESS.2021.3067499

[25] Md A. Uddin *et al*., “Data-driven strategies for digital native market segmentation using clustering,” *International Journal of Cognitive Computing in Engineering*, vol. 5, pp. 178-191, 2024, doi: 10.1016/j.ijcce.2024.04.002.

[26] X. Gong, X. Guo, R. Zhang, X. He, and A. Zhou, “Search Behavior Based Latent Semantic User Segmentation for Advertising Targeting,” in *2013 IEEE International Conference on Data Mining (ICDM)*, Dallas, TX, USA, 2013 pp. 211-220, doi: 10.1109/ICDM.2013.62.

[27] S. Sands, C. Ferraro, C. Campbell, and J. Pallant, “Segmenting multichannel consumers across search, purchase and after-sales,” *Journal of Retailing and Consumer Services*, vol. 33, pp. 62–71, 2016, doi: 10.1016/j.jretconser.2016.08.001.

[28] S. Kim, W. Shin, and H.-W. Kim, “Predicting online customer purchase: The integration of customer characteristics and browsing patterns,” *Decision Support Systems*, vol. 177, p. 114105, 2024, doi: 10.1016/j.dss.2023.114105.

BIOGRAPHIES OF AUTHORS

Sumit Kumar received a Bachelor's degree in Electronics and Telecommunication from Kurukshetra University, Kurukshetra, India, in 2005, a Master's degree from Guru Jambheshwar University of Science and Technology, Haryana, India, in 2008, and a Ph.D. degree from Jamia Millia Islamia, Delhi, India in 2017. Currently, he is working as a Professor at the Electronics and Telecommunication Department of Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, Maharashtra, India. His research areas are artificial intelligence, antenna design, IoT, wireless networks, wireless communication, and computational intelligence. He can be contacted at email: er.sumitkumar21@gmail.com.

Ruchi Rani received a Bachelor's degree in Computer Science Engineering from Kurukshetra University, Kurukshetra, India, in 2008, and a Master's degree from Maharshi Dayanand University, Haryana, India, in 2012. She is currently pursuing a Ph.D. degree with the Department of Computer Science Engineering, Indian Institute of Information Technology Kottayam, Kerala, India, and working as an Assistant professor at the Department of Computer Engineering and Technology, School of Computer Engineering and Technology, Dr. Vishwanath Karad MIT World Peace University, Pune, Maharashtra, India, Pune, India. Her research interests include machine learning and deep learning. She can be contacted at email: ruchiasija20@gmail.com.

Sanjeev Kumar Pippal is B.Tech. from MJP Rohilkhand University, M.Tech. and Ph.D. from MNNIT Allahabad, his area of interest is cloud computing, distributed computing and block chain. He is certified in machine and deep learning. He has published more than 30 research papers in SCI/Scopus international journals and conferences. He has filed four patents and published 02 patents. He was instrumental in setting up a 21 cr AI lab at AKTU, Lucknow. He has participated and coordinated as chair four BOS meetings at NIT Kurukshetra, Galgotias University and Chandigarh University. He has vast administrative experience as hod, dean, director at above mentioned organizations. He can be contacted at email: sanpippalin@gmail.com.

Riya Agrawal is pursuing her Bachelor of Technology in Computer Science Engineering at the Department of Computer Engineering and Technology, School of Computer Engineering and Technology, Dr. Vishwanath Karad MIT World Peace University, Pune, Maharashtra, India. Her research interests include machine learning and deep learning. She can be contacted at email: riyaagrawal1912@gmail.com.