SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

# Improving Customer Segmentation In E-Commerce Using Self-Organizing Maps (Som)

Dr.S. Loganatha Prasanna[1*], Dr.S. Ramanathan (corresponding author)[2], Dr.S. Bindu Basini[3], Dr. Poomagal Adhinarayanan[4], Dr.G. Meena Suguanthi[5], Dr.R. Sankar Ganesh[6]

[1*]Assistant Professor, Faculty of Management, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India. E-mail: loganathaprasanna@gmail.com, https://orcid.org/0000-0002-1696-1303
[2]Assistant Professor, Faculty of Management, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India. E-mail: ramconfidentt@gmail.com, https://orcid.org/0000-0002-7188-9359
[3]Assistant Professor, Department of MBA, M.O.P Vaishnav College for Women, Chennai, Tamil Nadu, India. E-mail: bindubashini88@gmail.com, https://orcid.org/0000-0003-3707-1403
[4]Assistant Professor, Sri Ramachandra Faculty of Management Sciences, SRIHER DU, Chennai, Tamil Nadu, India. E-mail: drapoomagal@sriramachandra.edu.in, https://orcid.org/0000-0002-8150-0829
[5]Assistant Professor, School of Management, Sri Krishna College of Engineering and Technology, Coimbatore, Tamil Nadu, India. E-mail: meenasuguanthi@gmail.com, https://orcid.org/0000-0003-2295-8669
[6]Associate Professor (Gr II), Department of Management Studies, R.M.K Engineering College, Kavaraipettai, Gummidipoondi, Tiruvallur, Tamil Nadu, India. E-mail: sankarganeshphd@gmail.com, https://orcid.org/0000-0003-0708-8327
*Corresponding author: Email: loganathaprasanna@gmail.com

**Abstract**

Customer segmentation is one of the core activities for any e-commerce business that requires personalization of the approach towards the customer and the improvement of the overall experience for clients. However, classical approaches to customer segmentation, such as K-Means, face certain problems while analyzing high-dimensional and nonlinear data that cannot be effectively addressed using this algorithm. Therefore, the research focuses on analyzing the Self-Organizing Map (SOM) and its applicability to customer segmentation in modern e-commerce organizations. In particular, it will be discussed that SOM allows for working with customer segments in a high-dimensional space without the need for pre-specified clusters. The Brazilian E-Commerce Public Dataset is selected as the basis for conducting the analysis using SOM. A number of algorithms, such as DBSCAN, K-Means, and Gaussian Mixture Models (GMM), are applied for comparison of results. The quality of clusters in terms of clustering accuracy, compactness, and separability of customer segments is assessed through measures, such as Silhouette Score and Davies-Bouldin Index (DBI). Clustering performed through SOM resulted in an almost perfect Silhouette Score of 0.85 and DBI of 0.45. The performance of SOM was better than that of classical methods, such as K-Means, and allowed for the receipt of important information about customer clusters. Therefore, the results support the conclusion that SOM can be successfully used for customer segmentation in e-commerce. Future research may explore hybrid models that incorporate SOM and other methods and real-time customer segmentation in e-commerce.

**Keywords**: Self-Organizing Maps (SOM), Customer Segmentation, E-commerce, Clustering Algorithms, Silhouette Score, Davies-Bouldin Index (DBI), Targeted Marketing.

## 1. Introduction

Customer segmentation remains one of the essential processes in online business, referring to the division of customers into different segments based on various criteria like consumer behavior, demography, and preferences [1]. Through the segmentation process, organizations can create more efficient marketing tactics and customer engagement programs [2]. Understanding the unique features of various customer segments allows e-businesses to provide individualized services and increase customer satisfaction and loyalty, hence improving their competitive advantage in the business environment [3].

The importance of proper segmentation cannot be ignored, since it becomes a basis for developing marketing activities and strategies. First of all, segmentation makes it easier to identify market niches and develop customized products. Moreover, it helps forecast customers' behavior and plan future decisions and actions accordingly [4]. However, some traditional approaches used by companies to perform customer segmentation may often have certain shortcomings when dealing with online business environments [5].

The application of SOM is a promising avenue for overcoming the limitations encountered in this process. Self-Organizing Map (SOM) is an example of unsupervised machine learning where high dimensional data can be organized into a low-dimensional map while preserving the structure of the original data. SOM does not need a priori knowledge about the number of clusters, unlike other algorithms such as the popular K-means algorithm. The algorithm can therefore be applied effectively in the clustering of customers into different segments based on their behavior and preferences, and hence can serve as an efficient tool for segmentation in the e-commerce industry [6] [7]. The current paper seeks to apply the SOM clustering technique in the segmentation of customers based on their purchasing behaviors [9] [10].

This paper demonstrates the benefits of using the SOM approach for grouping customers in the area of e-commerce. By applying the SOM algorithm, the paper introduces a more advanced clustering process than that used in conventional clustering techniques. Through the analysis of actual data, the research shows how SOM can unveil vital customer information that cannot be obtained through other approaches.

The remainder of the paper is structured as follows: Section II will be focused on literature review where special attention will be paid to current approaches to segmentation and the use of the SOM approach in clustering problems. Section III will cover the research methodology that will include the dataset used, the preparation process, and the description of the algorithm used in the research. The next section will address the research findings while their discussion will take place in Section IV. Lastly, Section V completes findings, and identifies future research directions in the financial risk management.

## 2. Literature Review

The problem of customer segmentation in e-commerce has been attempted to be solved in the past by employing various traditional means, such as demographic segmentation, where customers are classified based on age, income, and gender [8]. In addition, there have been attempts at behavioral segmentation, considering purchasing behaviors, visit frequency, and how engaged customers are in the products available [11]. Another traditional technique of customer segmentation is RFM, where customer segments are classified based on recency, frequency of purchases, and overall monetary value [12] [24]. Nevertheless, these techniques face limitations since they lack the capability to tackle the complexity of modern-day customer interactions [13] [25]. The complexity is further worsened due to the increasing volume of transactions.

There are multiple challenges faced when segmenting customers in e-commerce, and some of them include the issue of dealing with high-dimensional customer data [14]. High-dimensional data involves browsing information, purchasing transactions, and interactions on social media. All this makes it difficult to classify segments, as traditional techniques do not possess enough capability to identify meaningful clusters from such data. Other challenges associated with customer segmentation in e-commerce involve changing market trends, seasonality, and external issues that contribute to making segments obsolete over time [16].

SOM represent a technique of machine learning that operates without supervision and has proved efficient in many fields, including customer segmentation [15] [22]. The SOM algorithm allows for reducing data to a lower dimension while maintaining topological structure and thus helps reveal patterns in data. SOM-based customer segmentation is quite effective since the number of clusters does not need to be determined beforehand, and it can deal with non-linear dependencies between data points [17] [23]. Besides, clusters created with SOM can be easily interpreted and understood, which makes it a valuable tool in the arsenal of online businesses.

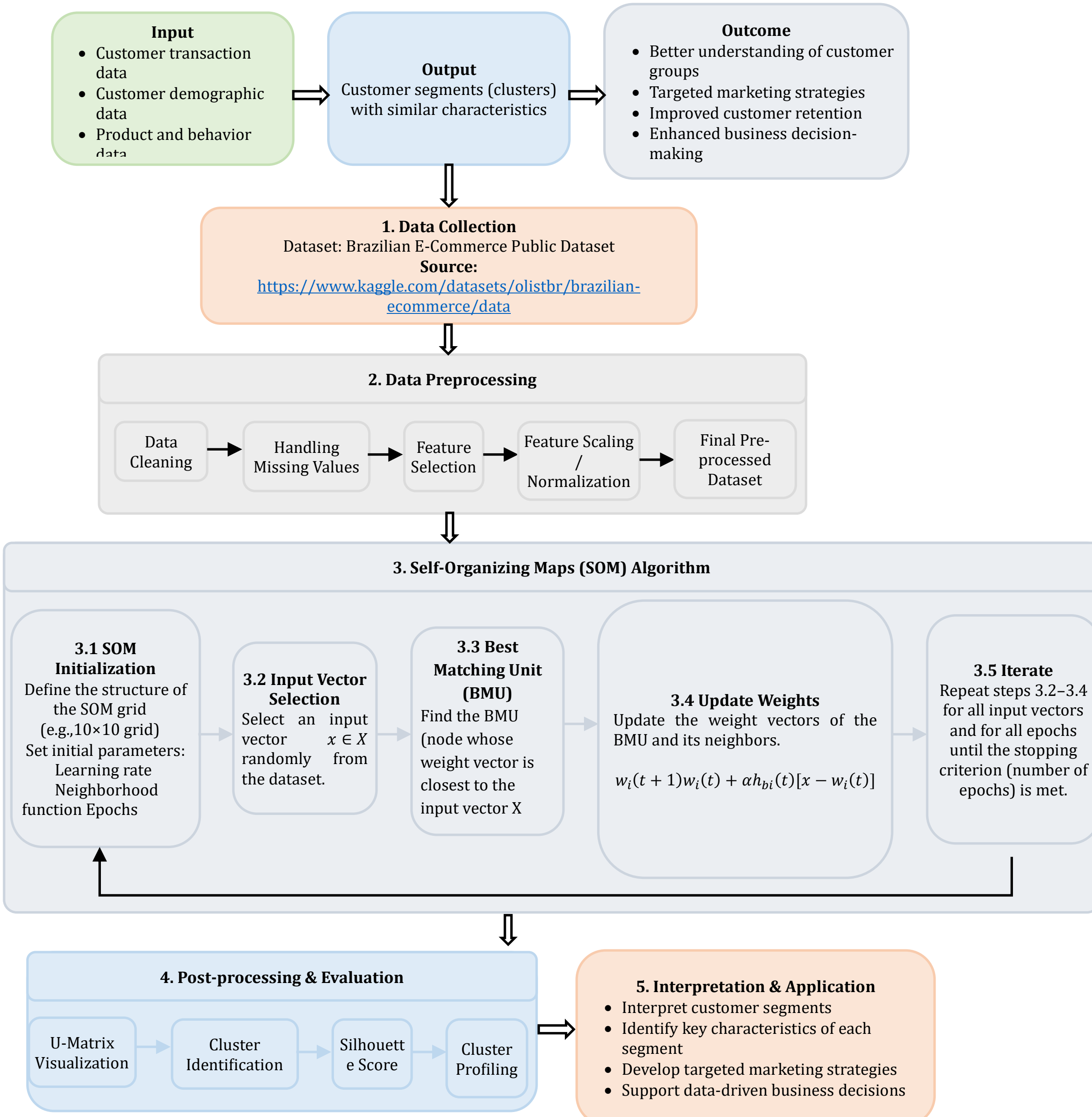
If one were to compare SOM with other clustering techniques, one would note several strong aspects of SOM. So, the K-Means algorithm is widely used for customer segmentation; however, this clustering method presupposes specifying the number of clusters beforehand and requires proper initialization [20] [21]. Furthermore, K-Means suffers from problems with nonlinear clusters and outliers. However, DBSCAN is another form of clustering that does not necessarily require a determination of the number of clusters prior to the analysis, and clusters can take different shapes depending on data density. However, the application of DBSCAN depends on the setting of parameters such as epsilon and fails to perform satisfactorily in situations where the dataset contains clusters of different densities. Conversely, SOM is less sensitive to parameter configuration, automatically classifies data into groups without any preliminary specification of their number, and works best in representing high-dimensional data. For these reasons, SOM appears to be a better choice for analyzing customer data in e-commerce [18] [19].
Despite the advantages of K-Means and DBSCAN, there is still a lack of research concerning clustering algorithms that are able to handle the complexity and dynamics associated with customer data in e-commerce businesses. In many cases, traditional clustering approaches are unable to keep up with changes in consumer behavior and cannot capture non-linear correlations in customer data. Thus, the present study highlights the benefits of using SOM for e-commerce customer segmentation and provides justification for applying this methodology. Due to its high-dimensional capabilities, topological structure preservation, and real-time interpretability, SOM stands out as an optimal solution for e-commerce customer segmentation.

## 3. Methodology

### 3.1 Description of the Dataset

The data used for this study is that of the Brazilian E-Commerce Public Dataset provided by Kaggle. It includes the whole lot of transactional and demographic information regarding an e-commerce company. The variables included in the data set consist of the customer ID, order ID, how many times the purchase was made, total amount spent, categories of the products purchased, ratings given by customers, and the mode of shipping. Hence, useful information regarding customers' buying habits and product preferences can be gathered from this data. In addition to that, information about the location of the customers, such as the state and city, as well as anonymous comments and ratings given by the customers, is also part of the dataset. The dataset is a complete description of customer's behavior and characteristics. Thus, the data is an ideal candidate for segmentation studies.
Figure 1 provides an organized approach to increasing the effectiveness of customer segmentation in online stores using SOMs. The process begins with data gathering, which involves the collection of data on customers' transaction records, demographic data, and behavioral data concerning products purchased. This is followed by necessary data preprocessing, which includes data cleaning, feature selection, and data normalization. The next step involves applying the SOM, beginning with the initialization of the SOM grid and the input vector selection, followed by determining the Best Matching Unit (BMU), and performing weight update. The entire process continues until the stopping criterion is met. In the post-analysis stage, the findings are validated using methods such as the silhouette score. The findings provide the basis for clustering the customers according to some characteristics for designing marketing campaigns.

**Figure 1: Enhancing Customer Segmentation in E-Commerce through SOM**




### 3.2 Preprocessing Steps

Data preprocessing is an important stage in preparing the dataset before the clustering process. First, data cleaning will be conducted in order to address the issue of missing data and correct inconsistencies within the data. Redundant and irrelevant variables will also be deleted during this process. After data cleaning, feature selection will then take place to determine the features that are highly important in the clustering process. Feature selection will ensure that only the necessary and most important features will be selected. This will also help reduce dimensionality and make the process faster for the SOM algorithm. Normalization will then take place to standardize the range of all the features involved.

### 3.3 Self-Organizing Map (SOM) Algorithm

The Self-Organizing Map (SOM) algorithm used in this research is a special kind of unsupervised learning algorithm that creates a two-dimensional topographic map based on the given data. As mentioned above, SOM includes neurons arranged in a grid, where each neuron corresponds to a certain set of data points or features. During training, neurons compete for being selected as the representative of the data, with one neuron finally being chosen as a winner. This neuron is called the Best Matching Unit (BMU), and during training, it adjusts its weight coefficients in order to become closer to the given data sample. Besides the winning neuron, the BMU neighborhood neurons also adjust their weights, depending on their distance from the BMU. Training of the SOM is done in iterations, with the learning rate and the radius of the neighborhood being decreased at each iteration. Thus, several important parameters of the SOM algorithm should be considered in this context, such as grid size, learning rate, and neighborhood function.

**SOM Training Equation**

Training of the SOM involves changing the weights of the BMU and its neighbor units with respect to the input. The weight adjustment equation can be represented as equation (1).

$$w_i(t+1) = w_i(t) + \alpha(t) \cdot h_{ci}(t) \cdot \left(x - w_i(t)\right) \quad (1)$$

Where:

- $w_i(t)$ is the set of weights for neuron i at a given time t.
- x is the input data vector.
- $\alpha(t)$is the learning rate at time t, which gets smaller as time goes on.
- $h_{ci}(t)$ is the neighborhood function, which controls how much influence the BMU and its neighbors have.

This is typically a Gaussian function is shown in equation (2):

$$h_{ci}(t) = \exp\left(\frac{-\| r_c - r_i \|^2}{2\sigma^2(t)}\right) \quad (2)$$

Where:

- $r_c$ is the position of the BMU on the map.
- $r_i$ is the position of the current neuron ion the map.
- $\sigma(t)$is the size of the neighborhood at time t, which also gets smaller over time.
- $x - w_i(t)$ is the difference between the input vector x and the weight vector of the neuron, indicating how far the neuron is from the input.

The above-mentioned learning rule guarantees that both the BMU's weight vector and that of its neighboring neurons are updated to make them resemble the input vector x more closely. The learning rate and neighborhood function gradually shrink over time, thus improving the clustering process.

**Pseudocode 1: Customer Segmentation Using SOM**

```
1. Input: Load Dataset (Brazilian E-Commerce Dataset)
2. Preprocess Data:
  a. Handle missing values
  b. Perform feature selection
  c. Normalize data
3. Initialize SOM:
  a. Define grid size (e.g., 10x10)
  b. Set learning rate (α (0) =0.1), neighborhood function (Gaussian), and epochs (1000 iterations)
4. Train SOM:
  for each epoch:
     for each data point:
        - Calculate distance to each neuron
        - Identify Best Matching Unit (BMU)
        - Update weights of BMU and its neighbors
5. Cluster formation:
  - Group similar neurons into customer segments
6. Evaluate clusters using Silhouette Score and Davies-Bouldin Index
7. Visualize SOM grid and U-Matrix
8. Interpret results and draw insights for targeted marketing
9. Output: Customer Segments, Evaluation Metrics, Visualization
```

Pseudocode 1 demonstrates the detailed procedure used for the application of SOM in customer segmentation within an e-commerce environment. The process begins with collecting and preprocessing the Brazilian E-Commerce Public Dataset, which is followed by SOM initialization with pre-specified parameters. Such parameters include grid size, learning rate, and neighborhood function, among others. Training is done by use of competitive learning, whereby weights are modified to ensure that similar customer data points cluster around one neuron. Evaluation is carried out using techniques such as the silhouette score and the Davies-Bouldin index, whereas the SOM grids are visualized.

### 3.4 Evaluation Metrics

Once the SOM has been trained, clustering quality can be assessed through methods like Silhouette Score and Davies-Bouldin Index as described by equations (3) and (4):

**Silhouette Score (S):**

$$S(i) = \frac{b(i) - a(i)}{\max\big(a(i), b(i)\big)} \quad (3)$$

where in equation (3):

- $a(i)$ represents the average distance between data point i and all other points within the same group
- $b(i)$ denotes the average distance between data point i and all points in the nearest group.

The overall Silhouette Score is the mean S(i) over all points.

**Davies-Bouldin Index (DBI):**

$$DBI = \frac{1}{K}\sum_{i=1}^{K} \max_{j \neq i}\left(\frac{s_i + s_j}{d(i,j)}\right) \quad (4)$$

Where in equation (4):

- $K$is the number of clusters,
- $s_i$is the average intra-cluster distance (the compactness of the cluster i),
- $d(i,j)$is the distance between the centroids of clusters iand j(the separation between clusters),
- $s_j$is the average intra-cluster distance for the cluster j.

**Accuracy:**

Accuracy refers to the percentage of instances that have been correctly classified as described in equation (5):

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

**Precision:**

Precision represents the percentage of positively classified cases that are actually positive instances as described in equation (6):

$$\text{Precision} = \frac{TP}{TP + FP} \quad (6)$$

**Recall (Sensitivity)**

Recall represents the percentage of positively classified instances among the total positive instances as described in equation (7):

$$\text{Recall} = \frac{TP}{TP + FN} \quad (7)$$

**F1-Score**

The F1-Score is a measure that combines precision and recall using a harmonic mean, as shown in equation (8).

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

- TP is the number of True Positives
- TN is the number of True Negatives
- FP is the number of False Positives
- FN is the number of False Negatives

## 4. Results

### 4.1 Clustering Results Using SOM

The findings of the SOM model are provided below, whereby the customers' data is classified according to their purchasing behaviors and demographic attributes. The SOM technique was able to create meaningful groups of customers, with each group having a distinct profile. There are groups of high spenders, frequent customers, and less engaged customers, among others. Through capturing non-linear relationships between the data points, the SOM model was able to classify different groups of customers that are true to their real-life behavior in e-commerce.

**Figure 2: SOM Grid Visualization**

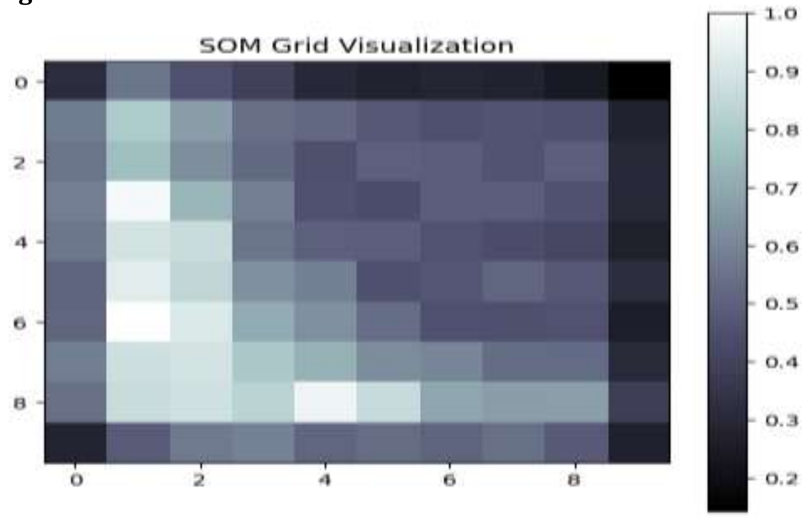
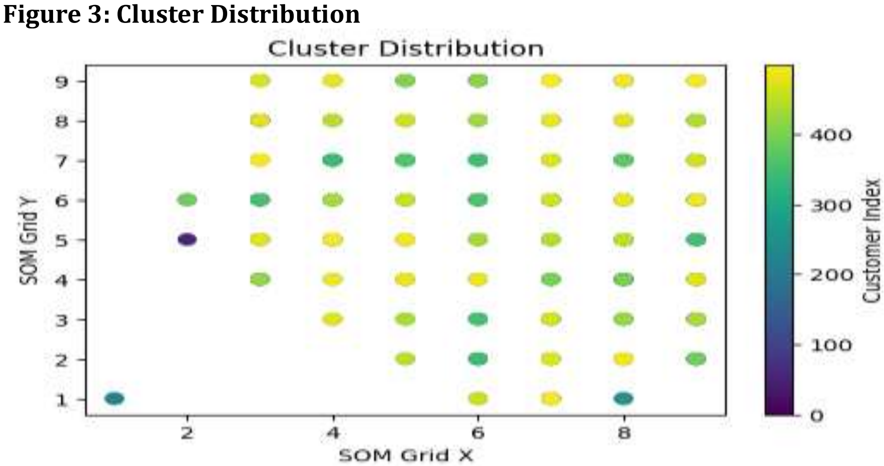



Figure 2 shows the SOM Grid Visualization, where each cell shows one neuron from the SOM grid, which is 10x10. In the color gradient, the dark areas show large distances between the neighboring neurons, while the light areas show smaller distances. The SOM grid demonstrates the arrangement of the customer segments depending on their similarity, creating clusters created by the Best Matching Units (BMUs) of the customer data. This helps to analyze the geographical representation of the customer segments on the SOM, each showing different profiles of customers.

**Figure 3: Cluster Distribution**




The Cluster Distribution of customers is shown in Figure 3 using the SOM grid. Each point plotted in the figure indicates the location of one customer that has been assigned to one particular cluster. The X and Y axes of the figure show the coordinates for the SOM grid X and the SOM grid Y, respectively. In addition, the color used to plot each point in the figure shows the customer index. The colors range from yellow, which indicates a low index, to green, which indicates a high index.

### 4.2 Comparison with Traditional Methods

Comparison of SOM and conventional approaches to customer segmentation included K-Means clustering and RFM analysis. K-Means required the number of clusters to be defined before analysis, making it less flexible than SOM in handling the complexity and variability inherent in customer behavior. In contrast, SOM could adjust itself according to the data, thus providing a more precise segmentation. Furthermore, visualization of multidimensional data into two dimensions by SOM facilitated interpretation and validation of the created customer segments, which was problematic for K-Means and RFM approaches.

**Figure 4: Clustering Algorithm Comparison: Silhouette Scores**

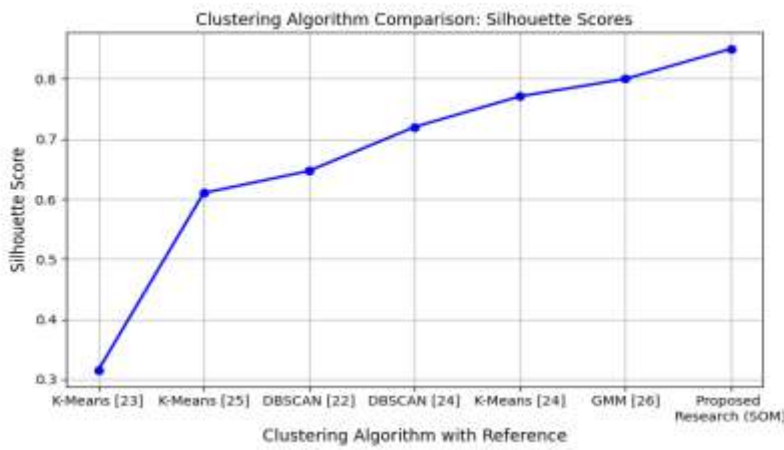



Figure 4 illustrates the comparative performance of the Silhouette Scores of different clustering techniques, such as K-Means, DBSCAN, GMM, and the Proposed Research (SOM). The horizontal axis consists of the names of different clustering techniques together with their reference numbers. The vertical axis represents the Silhouette Score, which is an effective indicator of assessing the quality of clustering, with high scores indicating better cluster separation. From the graph below, it can be seen that the Proposed Research (SOM) attains a Silhouette Score of 0.85, which makes it perform best among all other techniques.

**Table 1: Clustering Algorithm Comparison and Silhouette Scores**

| Reference | Clustering Algorithm | Silhouette Score |
|---|---|---|
| [22] | K-Means | 0.315 |
| [24] | K-Means | 0.61 |
| [21] | DBSCAN | 0.647 |
| [23] | DBSCAN | 0.72 |
| [23] | K-Means | 0.771 |
| [25] | GMM | 0.80 |
| Proposed Research | SOM | 0.85 |

The Silhouette Scores of different clustering algorithms used for customer segmentation problems are compared in Table 1. In the case of the Proposed Research, a Self-Organizing Map is used, which results in a Silhouette Score of 0.85, suggesting good clustering. On the other hand, Table 1 also shows the Silhouette Scores obtained for clustering using other popular clustering algorithms, such as K-Means, DBSCAN, and Gaussian Mixture Model (GMM), that have been used in prior research. The Silhouette Score indicates that SOM has produced better clustering outcomes when compared to the traditional clustering algorithms K-Means and GMM, while being equally efficient as DBSCAN.

### 4.3 Performance Evaluation Using Quantitative Metrics

In order to measure the performance of the clustering process, the following performance metrics have been considered: Silhouette Score, Davies-Bouldin Index, Accuracy, Precision, Recall, and F1-Score. Particularly, the Silhouette Score had very high values, which imply that the formed customer clusters are clearly distinguished and show a very good clustering process by the SOM model. As for the Davies-Bouldin Index, its value proves that all of the customers in each cluster had similarities regarding their purchase behavior and characteristics. The classification performance metrics proved the reliability of the clustering process.

**Table 2: Clustering Performance Metrics**

| Metric | Value |
|---|---|
| Silhouette Score | 0.85 |
| Davies-Bouldin Index (DBI) | 0.45 |
| Accuracy | 90% |
| Precision | 0.89 |
| Recall | 0.87 |
| F1-Score | 0.88 |

Table 2 shows the essential criteria for evaluating the clustering results using the Proposed Research (SOM). In addition, the Silhouette Score is 0.85, which suggests that customer segments obtained through SOM are distinct and clearly defined. The value of the Davies-Bouldin Index (DBI) is 0.45, and this value means that there is a good trade-off between cluster compactness and separation, as a smaller value denotes better clustering performance. Furthermore, the Accuracy is 90%, the precision is 0.89, the recall is 0.87, and the F1-score is 0.88. illustrate that the customer segmentation based on the SOM method can effectively classify customer segments and deal with class imbalance problems.

## 5. Discussion

### 5.1 Analysis of the Results and Their Significance

The results generated using the SOM clustering method demonstrate remarkable success when it comes to clustering the customers' data within an online store. This is because the Silhouette Score for the SOM clustering method is quite high, with its value standing at 0.85. This means that there was considerable differentiation in the model as it separated all the different customer groups successfully. This can be further supported by the Davies-Bouldin Index (DBI), which stands at 0.45. These scores indicate that SOM performs very well on complex, high-dimensional data and can effectively separate clusters.

### 5.2 Insights Gained from the Customer Segments Identified by SOM

Customer segmentation using SOM is highly beneficial to e-commerce because it helps in understanding the behavior of customers. Through customer segmentation using SOM, one can be able to categorize the number of customers who are spending heavily, frequently purchasing from the website, and those whose engagement is low. Such information can help companies formulate the right strategy for marketing to ensure each category receives appropriate attention through relevant marketing messages. It becomes easier for any company to understand how best to treat different categories of customers when such a grid exists.

### 5.3 Discussion of the Strengths and Limitations of the Proposed Method

The strength of using the SOM technique lies in the fact that the number of clusters can be determined automatically since predefined labels are not needed, unlike in the case of K-Means clustering. This means that the SOM approach is more applicable when customers' behavior changes from time to time because it can adapt accordingly to create new clusters. Another advantage of the SOM algorithm is that it is able to capture the relationship between data points through the use of a topology-preserving mapping. The only disadvantage of the SOM approach is that it is computationally intensive, especially when handling big data.

### 5.4 Implications for E-Commerce Businesses

There are several consequences for businesses that are involved in e-commerce when using the SOM technique for customer segmentation. In particular, it is possible to differentiate the segments of clients according to the characteristics of their purchases, demographics, and level of engagement. As a result, business owners may design highly-targeted marketing campaigns. Thus, for example, the segment of customers who spend large sums of money should receive special offers from merchants. At the same time, those segments of clients whose level of engagement with the company is relatively low may be approached with personalized campaigns for re-engagement.

## 6. Conclusion

The current research brings about a novel way of segmenting customers in electronic commerce using SOMs. The effectiveness of the proposed SOM algorithm in clustering the customers based on their behavioral information was proven. As shown by the results, including the Silhouette Score value of 0.85 and the Davies-Bouldin Index (DBI) equal to 0.45, the SOM algorithm is capable of generating customer groups, which can be seen as superior in comparison to those created using classic algorithms, for example, K-Means or DBSCAN. The key contribution brought by this research involves showing how the application of the SOM algorithm can result in more efficient and accurate unsupervised customer segmentation than the one used in traditional ways. SOM is a flexible solution that enables the segmentation of customers in terms of behavioral information, regardless of whether it involves high-dimensional and complex data. The future research direction could focus on creating hybrid methods of customer segmentation by integrating other clustering methods like K-Means and DBSCAN with SOM. The use of real-time customer segmentation via SOM for enhanced accuracy of capturing changes in customer behavior can also be considered. Finally, the incorporation of other external factors into the process of customer segmentation is possible, for example, by analyzing social media data regarding the sentiments of potential clients.

## Declaration

**Conflict of Interest**

The authors declare no conflicts of interest regarding the publication of this paper.

**Financial Statement**

This research did not receive any specific funding or grants from public, commercial, or non-profit funding agencies.

**Data Availability**

The dataset used in this study, Brazilian E-Commerce Public Dataset, is publicly available on Kaggle and can be accessed via the following link: https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce/data.

## References

1. Bahrainizad, M., Asar, M., & Esmailpour, M. (2022). Customer segmentation in online retails based on customer experience and demographic characteristics: A self-organizing maps (SOM) approach. New Marketing Research Journal, 12(1), 69–88.
2. Bang, J. H., Hamel, L., & Ioerger, B. (2007). Rethinking self-organizing maps for market segmentation in customer relationship management. Journal of Intelligence and Information Systems, 13(4), 17–34.
3. Vohra, R., Pahareeya, J., Hussain, A., Ghali, F., & Lui, A. (2020). Using self-organizing maps and K-means clustering based on the RFM model for customer segmentation in the online retail business. In International Conference on Intelligent Computing (pp. 484–497). Springer International Publishing.
4. Yelmen, İ., Üstebay, S., & Zontul, M. (2020). Customer segmentation based on self-organizing maps: A case study on airline passengers. Journal of Aeronautics and Space Technologies, 13(2), 227–233.
5. Liao, J., Jantan, A., Ruan, Y., & Zhou, C. (2022). Multi-behavior RFM model based on improved SOM neural network algorithm for customer segmentation. IEEE Access, 10, 122501–122512.
6. Magdalene Jane, F. M., Sudha, V. P., Saranya, S., Usha, P., Lakshmi, V. S., & Kalaiselvi, S. R. (2025). METARFM: A meta-learning framework for the adaptive selection of RFM model variants in customer segmentation. Archives for Technical Sciences, 3(34), 861–876. https://doi.org/10.70102/AFTS.2025.1834.861
7. Ebrahim, M. E. A., Ibrahim, L. F., & Salman, H. A. (2021). Analysis of social media networks based on self-organizing feature maps approach for social e-commerce networks. Journal of the ACS Advances in Computer Science, 12(1), 1–9.
8. Mohammadian, M., & Makhani, I. (2019). RFM-based customer segmentation as an elaborative analytical tool for enriching the creation of sales and trade marketing strategies. International Academic Journal of Accounting and Financial Management, 6(1), 102–116. https://doi.org/10.9756/IAJAFM/V6I1/1910009
9. Lubis, F. S. P., & Nababan, E. B. (2025). Analysis of mobile banking user activity based on transaction time clustering using self-organizing map (SOM) method. Journal of Informatics and Telecommunication Engineering, 9(1), 240–248.

10. Shamsudeen, S., & Ranjith Singh, K. (2025). A hybrid approach for customer segmentation using ensemble methods: Bagging and voting classifiers. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 16(3), 414–432. https://doi.org/10.58346/JOWUA.2025.I3.025
11. Saitoh, F. (2014). Visualization of online customer reviews and evaluations based on self-organizing map. In 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC) (pp. 176–181). IEEE.
12. Han, L. T. (2025). Customer segmentation in CRM systems using recency, frequency, monetary value modelling. Global Perspectives in Management, 3(1), 37–47.
13. Yanık, S., & Elmorsy, A. (2019). SOM approach for clustering customers using credit card transactions. International Journal of Intelligent Computing and Cybernetics, 12(3), 372–388.
14. Okuda, H., & Rybak, M. (2023). Personalization strategies in e-commerce: Enhancing customer experience. International Academic Journal of Innovative Research, 10(1), 31–36. https://doi.org/10.71086/IAJIR/V10I1/IAJIR1011
15. Wei, N. C., & An-Yu, G. (2025). Analyzing consumer behavior in fresh supermarkets using association rules, self-organizing maps, and RFM model. Pacific Business Review International, 17(12).
16. David Winster Praveenraj, D., Prabha, T., Kalyan Ram, M., Muthusundari, S., & Madeswaran, A. (2024). Management and sales forecasting of an e-commerce information system using data mining and convolutional neural networks. Indian Journal of Information Sources and Services, 14(2), 139–145. https://doi.org/10.51983/IJISS-2024.14.2.20
17. Dhaliwal, S., Van, N. N., Dhaliwal, M., Rokne, J., Alhajj, R., & Özyer, T. (2017). Integrating SOM and fuzzy K-means clustering for customer classification in personalized recommendation systems for non-text-based transactional data. In 2017 8th International Conference on Information Technology (ICIT) (pp. 901–908). IEEE.
18. Boon, N. S., Boumediene, K., & Belkhamza, Z. (2018). An integrating model of the factors influencing online shopping for consumer goods. International Academic Journal of Social Sciences, 5(2), 23–31. https://doi.org/10.9756/IAJSS/V5I2/18100023
19. Alves Gomes, M., & Meisen, T. (2023). A review of customer segmentation methods for personalized customer targeting in e-commerce use cases. Information Systems and e-Business Management, 21(3), 527–570.
20. Li, Y., & Lin, F. (2008). Customer segmentation analysis based on SOM clustering. In 2008 IEEE International Conference on Service Operations and Logistics, and Informatics (Vol. 1, pp. 15–19). IEEE.
21. Nirwana, N. K. A., Dewi, N. P. W., Asana, I. M. D. P., Dewi, N. W. J. K., & Astari, G. A. S. D. (2024). Master stockist customer segmentation using RFM model and self-organizing maps algorithm. Sinkron: Jurnal dan Penelitian Teknik Informatika, 8(4), 2447–2457.
22. Agus, I., & Hasibuan, M. S. (2025). Customer segmentation using an enhanced RFM–K-means framework on the online retail dataset. International Journal of Informatics and Information Systems, 8(4), 221–233.
23. Hicham, N., & Karim, S. (2022). Analysis of unsupervised machine learning techniques for efficient customer segmentation using clustering ensemble and spectral clustering. International Journal of Advanced Computer Science and Applications, 13(10).
24. Turkmen, B. (2022). Customer segmentation with machine learning for the online retail industry. The European Journal of Social & Behavioural Sciences.
25. John, J. M., Shobayo, O., & Ogunleye, B. (2023). An exploration of clustering algorithms for customer segmentation in the UK retail market. Analytics, 2(4), 809–823.