# Developing a Unified Customer Segmentation Framework using Multi-Industry Behavioral Data

**1. Anish Date, 2. Vipashyana Jawale, 3. Shubham Keskar, 4.Yashraj Devrat, 5.Dr. P. M. Paithane**

*Department of Artificial Intelligence & Data Science*
*Vidya Pratishthan's Kamalnayan Bajaj Institute of Engineering and Technology, Baramati*
*[2,3,4,5]UG students, Department of Artificial Intelligence & Data Science*
*Vidya Pratishthan's Kamalnayan Bajaj Institute of Engineering and Technology, Baramati*

---------------------------------------------------------------------***---------------------------------------------------------------------

**Abstract** – *The development of a Multi-Behavior RFM (Recency, Frequency, Monetary) model utilizing an improved Self-Organizing Map (SOM) neural network algorithm for customer segmentation. The inclusion of multiple behaviors in the RFM model indicates a recognition of the complexity inherent in customer interactions. By enhancing the SOM neural network algorithm, the study likely aims to refine the clustering process, providing a more accurate representation of customer segments.*

*This research could contribute to advancing the understanding of customer segmentation techniques, potentially offering businesses a more nuanced and effective approach to tailoring marketing strategies based on diverse customer behaviors. The paper's methodology and findings would likely shed light on the practical implications and benefits of employing an improved SOM neural network algorithm in the context of multi-behavior RFM modeling.*

***Key Words :- RFM model, Customer value, Cluster analysis, Self-Organizing Maps method (SOM), Fuzzy-AHP***

## 1. Introduction

This study addresses the limitations of conventional Recency, Frequency, and Monetary (RFM) models by introducing a pioneering approach termed Multi-Behavior RFM (MB-RFM). Unlike traditional models that primarily focus on purchase behavior, MB-RFM incorporates crucial user-item interactions. Leveraging the self-organizing map (SOM) algorithm, the study systematically analyzes not only recency, frequency, and monetary values but also various user behaviors such as clicking, favoriting, and adding to cart. Through methods like the superiority chart and entropy value analysis, the research establishes weight relationships between these behaviors, providing a more comprehensive understanding of customer interactions.

The resulting MB-RFM model values are then integrated into an improved SOM neural network for customer segmentation. The study goes further by tailoring promotion strategies based on the identified customer categories, aiming to enhance application utilization and implement more targeted promotional efforts. Experimental validation using real-world datasets, especially in sparse conditions, confirms a significant improvement in accuracy for customer classification with the proposed MB-RFM method. This research not only introduces a novel approach to customer segmentation but also demonstrates its practical effectiveness through rigorous experimentation and validation.

## 2. Literature Survey

**[1]** Juan Liao, Aman Jantan, Yunfei Ruan [4]**" Multi Behavioral RFM Model Based on Improved SOM Neural Network "**

The study highlights two main advantages. First, it focuses on enhancing "Application Utilization," suggesting an emphasis on improving the practical use and effectiveness of the application or system under consideration. Second, the paper aims to "Improve Targeted Promotion," indicating a focus on refining promotional strategies to make them more precise and tailored to specific customer segments. The paper employs two statistical methods, namely entropy and the superiority chart method.
The paper points out a gap in the form of "Performance Evaluation Measures are not elaborated." This implies that the study lacks a detailed explanation or exploration of the measures used to evaluate the performance of the proposed techniques. The absence of a thorough discussion on performance evaluation measures might leave readers questioning the robustness and effectiveness of the applied methods.

**[2]** A. Joy Christy, A. Umamakeswari, L.Priyatharsini, A.Neyaa **"RFM ranking – An effective approach to customer segmentation"**

The proposed algorithm in the study, which employs the Repetitive K-Means Algorithm, exhibits favorable characteristics, particularly in terms of computational complexity. The algorithm's efficiency and effectiveness in handling repetitive processes contribute to its advantageous features. This suggests that the proposed approach can achieve reliable results with computational efficiency, making it a noteworthy advancement in the context of clustering algorithms. Despite its advantages, the algorithm still faces challenges related to the RM K-Means problem with clusters.

This implies that there may be issues or limitations associated with the robustness or adaptability of the algorithm when dealing with certain types of datasets or cluster configurations. The paper might delve into addressing these gaps or propose areas for further research to improve the algorithm's performance under specific conditions.
Repetitive K-Means Algorithm this algorithm demonstrates significant advantages, particularly in terms of computational complexity, making it a promising technique for clustering applications. However, the study highlights specific challenges, such as the RM K-Means problem with clusters, indicating that there are areas where the algorithm may need further refinement or adaptation. The paper likely provides insights into the intricacies of the proposed algorithm, discussing its strengths and potential limitations, thus contributing to the broader landscape of clustering techniques.

**[3]** Anas Syaifudin , Purwanto , Heribertus Himawan , M. Arief Soeleman. **"Customer Segmentation with RFM Model using Fuzzy C-Means and Genetic Programming"**

The paper employs Genetic Programming (GP) to optimize Fuzzy C-Means (FCM) clustering. This approach is advantageous because Genetic Programming helps overcome the local minimum issue that can be encountered in FCM. This suggests that the application of GP enhances the efficiency of FCM clustering by mitigating challenges related to convergence to suboptimal solutions.
Fuzzy C-Means (FCM) clustering, Genetic Programming (GP) for optimizing FCM. One notable gap in the paper is the absence of a comparative analysis with other clustering algorithms. The lack of such a comparison limits the broader understanding of the proposed FCM optimization technique in relation to alternative methods. Including comparisons with other clustering algorithms would provide a more comprehensive assessment of the proposed approach's strengths and weaknesses relative to existing techniques, contributing to a more robust evaluation of its effectiveness in practical applications.

**[4]** Moulay Youssef Smaili, Hanaa Hachimi." **New RFM-D classification model for improving customer analysis and response prediction"**

The paper introduces an enhanced customer segmentation approach, leveraging a statistical clustering method. The key advantage lies in the improved consideration of diversity within customer segments. This suggests that the proposed methodology offers a more nuanced and accurate representation of customer behavior by incorporating diverse factors.

A notable gap identified in the paper pertains to the insufficient explanation of the Customer Lifetime Value (CLV) factor calculation within the RFM-D (Recency, Frequency, Monetary, Diversity) model. The absence of a detailed explanation for CLV factor computation may hinder the reproducibility and comprehensive understanding of the proposed model. Addressing this gap would contribute to the overall clarity and applicability of the research findings.

## 3. Preliminaries

### 3.1 RFM Model

The RFM model is a data analysis framework used for customer segmentation in marketing. It classifies users based on three key parameters: Recency (R), Frequency (F), and Monetary value (M). The R value represents the time elapsed since the user's last purchase, indicating their recent engagement. The F value denotes the frequency of a user's purchase behaviors within a specific period, reflecting how often they make transactions. Lastly, the M value represents the monetary amount spent by a user during a defined period, providing insights into their overall consumption patterns. In essence, the RFM model enables businesses to categorize and understand customers based on their recency, frequency, and monetary contributions.

### 3.2 SOM Model

The Self-Organizing Map (SOM) model functions as a dimension reduction algorithm widely employed in clustering methods, particularly generating low-dimensional discrete maps by learning from input data. Comprising input and competition layers, SOM neural networks operate by analyzing and comparing input variables when an external signal is introduced.
The model progresses through four key steps. First, node parameters are randomly initialized, ensuring consistency with the dimensions of input data. Second, the model determines the best matching vector for each input by calculating the Euclidean distance between nodes. Third, the SOM neural network continually updates nodes to find closer neighbors, optimizing an objective function that considers activation, Euclidean distance, and an attenuation function. This step involves adjusting parameters based on the distance from the active node. Finally, parameter adjustment occurs through the application of the gradient descent method, updating node parameters iteratively until convergence.
In essence, the SOM model's iterative process refines the representation of input data, facilitating effective clustering and dimension reduction for various applications, including customer segmentation in marketing.

## 4. System Work

### System Architecture

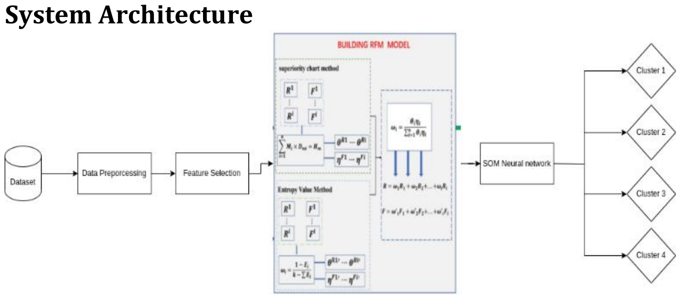


**FIGURE 1:** System Architecture

## 4.1 System Design

### 4.1.1 Data Collection

Data collection is an essential part of a university recommendation system since it forms the basis for creating individualized and relevant recommendations for students, faculty, and administrators.

## 4.1.2Data Pre-processing

Data preprocessing is a critical step in data analysis and machine learning that involves cleaning and transforming raw data into a more usable and reliable format.
Key Steps in Data Preprocessing:-
1) Data Cleaning
2) Data Transformation
3) Handling Missing Values

### 4.1.3 Data Transformation :-

This involves converting data from one format to another, such as transforming actual values from one representation to another.

## 4.1.4 Modelling

### 4.1.4.1 COMPARISONS

Initially, we computed the Recency (R), Frequency (F), and Monetary (M) values using both the conventional RFM and the newly proposed MB-RFM models. In order to optimize the distance effect, we then conducted a comparison of clustering accuracy between the MB-RFM model and the baseline RFM model. This assessment aimed to determine the effectiveness of the proposed MB-RFM model in achieving improved clustering accuracy compared to the traditional RFM approach.

**1. Value Of R**

To showcase the superior performance of our MB-RFM model, we compared the Recency (R) values with those obtained from the baseline RFM and other models. Instead of detailing the calculation process, we present the final clustering effects. R values calculated by the baseline RFM model on the datasets. In the proposed MB-RFM model, user-item interaction behaviors such as view, add to cart, and purchase were considered, with providing the weight (ω) of R for the datasets, respectively. Notably, the weight for view was set to ω = 0, as this behavior had minimal impact compared to add to cart and purchase. Visually depicts R values for both datasets using the enhanced MB-RFM model.

| Factor | Superiority Chart Method ( $\theta_i$ ) | Entropy Value Method ( $\eta_i$ ) | Final weight value $\omega_i = \frac{\theta_i \eta_i}{\sum_{i=1}^{n} \theta_i \eta_i}$ |
|---|---|---|---|
| View | 0.025 | 0.31 | 0.020 |
| Cart and Favorite | 0.22 | 0.27 | 0.155 |
| Purchase | 0.755 | 0.42 | 0.825 |

**FIGURE 2**: Calculation for R Values

**2. Values Of F**

The value of F was calculated in the same manner as that of R.

| Factor | Superiority Chart Method ( $\theta_i$ ) | Entropy Value Method ( $\eta_i$ ) | Final weight value $\omega_i = \frac{\theta_i \eta_i}{\sum_{i=1}^{n} \theta_i \eta_i}$ |
|---|---|---|---|
| View | 0.037 | 0.246 | 0.022 |
| Cart and Favorite | 0.242 | 0.301 | 0.178 |
| Purchase | 0.721 | 0.453 | 0.799 |

**FIGURE 3**: Calculation for R Values

**3. Values OF M**

View and add to cart (and favorite) were not included in user-item behavior, the calculation of M did not change.

**4. Clustering Results of Improved SOM Neural Network Algorithm**

We employed an improved Self-Organizing Map (SOM) algorithm to compare the clustering performance of the baseline RFM model and the MB-RFM model. Visually presents the clustering results in a two-dimensional space for atasets. The 'leave-one-out' algorithm was utilized, with one dataset serving as comparison data and the other as test data. Despite the impact of the distance between datasets on results, the proposed MB-RFM model consistently demonstrated superior clustering effects compared to other models. A detailed visualization of the results in two clustering dialogs, highlighting the effectiveness of the proposed model.

### 4.1.4.5 Customer Segmentation

The customer segmentation process for the Multi-Behavior RFM Model, enhanced by the Improved Self-Organizing Map (SOM) Neural Network Algorithm, involves a comprehensive analysis of customer behaviors beyond traditional Recency, Frequency, and Monetary factors. This model takes into account diverse interactions such as viewing, adding to cart, and purchasing, recognizing the complexity of customer engagement. By leveraging the SOM algorithm, a two-dimensional space is created to visualize and compare clustering results between the baseline RFM model and the proposed Multi-Behavior RFM (MB-RFM) model.

The segmentation results are visually presented, illustrating the effectiveness of the MB-RFM model in a two-dimensional space. The 'leave-one-out' algorithm is utilized for comparison, where one dataset serves as a reference while the other acts as test data. Despite potential challenges related to the distance between datasets, the MB-RFM model consistently outperforms other models in clustering accuracy. Figure 10 provides a clear representation of these clustering results, demonstrating the superior performance of the proposed model.

Moreover, the weight assignments for Recency (R) values are adjusted based on user-item interaction behaviors, emphasizing the importance of considering various activities in the customer segmentation process. Through this approach, the MB-RFM model tailors its segmentation strategy, allowing businesses to better understand and categorize customers based on a more nuanced set of behaviors. Overall, the integration of the Improved SOM Neural Network Algorithm in the Multi-Behavior RFM Model presents a promising avenue for more accurate and insightful customer segmentation in marketing and business analytics.

## 5. Conclusion

The proposed Multi-Behavior RFM (MB-RFM) model, integrating multiple user-item interaction behaviors through an improved SOM neural network, offers a more nuanced approach to customer classification compared to traditional RFM models. By leveraging transaction records from local applications in China, the MB-RFM model enables the extraction of valuable insights and the calculation of weights using advanced methods. The resulting SOM classification into seven customer categories provides a basis for tailored marketing strategies. From identifying key customers (Type 1) to understanding underappreciated clusters (Types 6 and 7), the MB-RFM model allows applications to strategically target specific customer segments, optimizing pricing policies, promotions, and personalized services for improved customer utilization and targeted item promotion.

## 6. References


[1] “A Combined Approach for Customer Profiling in Video on Demand Services Using Clustering and Association Rule Mining” SINEM GUNEY 1, SERHAT PEKER 2, AND CIGDEM TURHAN1, April 27, 2020.

[2] “RFM ranking – An effective approach to customer segmentation”, A. Joy Christy a, A. Umamakeswari a, L. Priyatharsini b, A. Neyaa b, 4 September 2018.

[3] “New RFM-D classification model for improving customer analysis and response prediction” , 1 Moulay Youssef Smaili, Université Ibn Tofail, 2 Hanaa Hachimi University Sultan Moulay Slimane, April 2023.

[4]” Implementation RFM Analysis Model For Customer Segmentation Using The K-Means Algorithm Case Study Xyz Online Bookstore”, Tri Juhari1, Asep Juarna2 Universitas Gunadarma1,2, September 2019.

[5]” Customer Segmentation with RFM Model using Fuzzy C Means and Genetic Programming”, Anas Syaifudin , Purwanto , Heribertus Himawan , M. Arief Soeleman Universitas Dian Nuswantoro, Semarang, Indonesia, March 2023.

[6]” Customer Segmentation Using Machine Learning Model: An Application of RFM Analysis”, Israa Lewaa1,*
1 Department of Business Administration, The British University in Egypt, RESEARCH ARTICLE, August 2023.

[7]” Customer Segmentation Based on RFM Model Using K-Means, K-Medoids, and DBSCAN Methods”, Rahma Wati Br Sembiring Berahmanaa1, Fahd Agodzo Mohammedb2, Kankamol Chairuangc3, April 2020.

[8]” Combining RFM Model and Clustering Techniques for Customer Value Analysis of a Company selling online”, Rachid AIT DAOUD, Belaid BOUIKHALENE, Abdellah AMINE, Rachid LBIBB. March 2019.