Journal of Posthumanism
2024
Volume: 4, No: 3, pp. 964–996
ISSN: 2634-3576 (Print) | ISSN 2634-3584 (Online)
posthumanism.co.uk

DOI: https://doi.org/10.63332/joph.v4i3.424

Fuzzy Clustering Approach to Consumer Behavior Analysis Based on Purchasing Patterns

Anber Abraheem Shlash Mohammad[1],Yogeesh N[2*],Suleiman Ibrahim Shelash Mohammad[3,4], N Raja[5], Lingaraju[6], P. William[7], Asokan Vasudevan[8], Mohammad Faleh Ahmmad Hunitie[9]

Abstract

Consumer behaviour analysis is critically important to contemporary marketing strategy, allowing for an understanding of purchasing patterns and design of targeted interventions. Traditional segmentation methods like K-means and hierarchical clustering are often not sufficient as they force consumers into a single cluster, though life is not set in stone and real worldbehaviour is continuously overlapping. This research utilizes fuzzy clustering (machine learning) to segment consumers across Product Category Preference (PCP), Price Sensitivity (PS) and Purchase Frequency (PF) using the Fuzzy C-Means (FCM) algorithm. The report identifies two underlying consumer segments: High Spend Customers, who have strong product preference, purchase often, and are not sensitive to price, and Bargain Shoppers who are price-sensitive and make infrequent purchases with low product preference. Fuzzy clustering assigns each customer as a member to each of the segments with a value between 0 and 1, rather than in a traditional, discrete method of forcing segmentation, capturing the flexible, context-dependent nature of purchasing behaviour. In this process, the model iterates over the computation of cluster centroids and updates cluster membership until it reaches a stable state. The study finds that fuzzy clustering is more representative of the hybrid and uncertain nature of consumer behaviour which provides businesses with the benefit of adapting marketing strategies accordingly. Beyond the segmentation of consumers, fuzzy clustering may also find application in personalization, recommendation engines (including recommendation and personalization of products), dynamic pricing (where customers are targeted on special offers), loyalty programs, and optimization of the retail space (to better serve customers using both traditional and modern selling formats).

Keywords: *Fuzzy Clustering, Customer Segmentation, Consumer Behaviour Analysis, Fuzzy C-Means Algorithm, Membership Degree, Bargain Shoppers, High Spend Customers, Overlapping Clusters, Purchase Frequency, Product Preference, Price Sensitivity, Marketing Analytics.*

[1]Digital Marketing Department, Faculty of Administrative and Financial Sciences, Petra University, Jordan. Email: mohammad197119@yahoo.com
[2*]Department of Mathematics, Government First Grade College, Tumkur, Karnataka, India. Email: yogeesh.r@gmail.com
[3] Electronic Marketing and Social Media, Economic and Administrative Sciences Zarqa University, Jordan.
[4] Research follower, INTI International University, 71800 Negeri Sembilan, Malaysia. Email: dr_sliman@yahoo.com ORCID: (0000-0001-6156-9063)
[5]Assistant Professor, Sathyabama Institute of Science and Technology, Department of Visual Communication, Chennai, Tamil Nadu.
[6] Department of Physics, Government First Grade College of Arts, Science and Commerce, Sira, Tumkur, Karnataka, India.
[7] Department of Information Technology, Sanjivani College of Engineering, Savitribai Phule Pune University, Pune, India
[8] Faculty of Business and Communications, INTI International University, 71800 Negeri Sembilan, Malaysia.
[9] Department of Public Administration, School of Business, University of Jordan, Jordan Email: Mhunitie57@gmail.com

Introduction

Overview of Consumer Behaviour Analysis and Its Importance in Understanding Purchasing Patterns

Consumer behaviour analysis is the study of how individuals, groups, or organizations select, purchase, and use products, services, or ideas to satisfy their needs and desires. Understanding consumer behaviour is crucial for businesses to align their products and marketing strategies with customer preferences (Kotler & Keller, 2016; Mohammad et al., 2024a; Mohammad et al., 2024b). Purchasing patterns reveal consumer tendencies such as frequency of purchases, preferred product categories, and price sensitivity.

Consumer preferences can be quantified using feature variables such as:

- **Product Category Preference (PCP)**: The degree to which a customer prefers a particular product category.
- **Price Sensitivity (PS)**: The responsiveness of a customer to price changes.
- **Purchase Frequency (PF)**: The frequency of purchasing behaviour.

These features form a **multidimensional data point** for each consumer:

$$x_i = [\mathrm{PCP}_i, \mathrm{PS}_i, \mathrm{PF}_i]^T \in \mathbb{R}^3$$

Where x_i represents the feature vector for customer i.

Consumer behavior analysis involves grouping similar consumers based on these variables, enabling businesses to create personalized marketing strategies and optimize inventory and pricing policies (Solomon, 2018; Mohammad et al., 2024c).

Significance of Customer Segmentation in Developing Targeted Marketing Strategies

Customer segmentation is the process of dividing a broad customer base into smaller groups with shared characteristics. Segmentation allows firms to focus resources on the most promising market segments, leading to improved customer satisfaction and business performance (Wedel & Kamakura, 2000; Mohammad et al., 2024d).

Mathematically, segmentation can be expressed as:

Given a dataset $X = \{x_1, x_2, \dots, x_n\}$ with n customers and each customer represented as a feature vector in $\mathbb{R}^d$, the goal is to partition X into C distinct segments:

$$X = \bigcup_{c=1}^{C} X_c, \quad X_c \subseteq X, \quad X_i \cap X_j = \emptyset \text{ for } i \neq j$$

This partitioning is optimized to minimize the intra-cluster variance and maximize inter-cluster distance.

Targeted marketing strategies leverage segmentation results to deliver relevant promotions, pricing, and product recommendations (Dolnicar et al., 2018; Mohammad et al., 2024e).

Introduction to Fuzzy Clustering as a Method Capable of Handling Overlapping Consumer Preferences

Traditional clustering methods, such as K-means, assign each data point to a single cluster, assuming that each customer belongs exclusively to one segment. However, consumer behaviour is often overlapping and uncertain (Bezdek, 1981; Mohammad et al., 2024f).

Fuzzy clustering provides a more flexible and realistic approach by allowing customers to belong to multiple segments with varying degrees of membership.

Mathematical Definition of Fuzzy Partition Matrix: Let $U \in [0,1]^{n \times C}$ represent the fuzzy membership matrix, where:

$$u_{ic} \in [0,1], \quad \sum_{c=1}^{C} u_{ic} = 1, \quad \forall i = 1,2, \dots, n$$

u_{ic} is the membership degree of customer i in cluster c.

Partial membership captures the consumer's tendency towards multiple segments, e.g., sometimes being a high spender and other times seeking discounts.

Objectives of the Study

This study aims to:

- Segment consumers based on their product preferences, price sensitivity, and purchase frequency.
- Apply fuzzy clustering to model overlapping customer behaviour.
- Evaluate the convergence and stability of the clustering algorithm using membership degree updates and centroid calculations.
- Interpret the results to derive actionable business insights.

Literature Review

Understanding consumer behavior patterns is essential for businesses to improve customer targeting and optimize marketing strategies. Customer segmentation plays a critical role in grouping similar consumers based on their purchasing behavior. Over the years, various clustering techniques have been employed in consumer behavior analysis, with K-means and hierarchical clustering being the most popular.

2.1. Traditional Clustering Techniques in Consumer Segmentation

Traditional clustering methods assign each data point to exactly one cluster.

K-means clustering (MacQueen, 1967; Mohammad et al., 2024g) minimizes the within-cluster sum of squared distances:

$$J = \sum_{i=1}^{n} \sum_{c=1}^{C} \delta_{ic} \|x_i - v_c\|^2$$

Where:

v_c is the centroid of cluster c.

$\delta_{ic} = 1$ if customer i is assigned to cluster c, otherwise $\delta_{ic} = 0$.

Hierarchical clustering builds a tree-like structure based on data similarity, merging or splitting clusters iteratively (Johnson, 1967; Mohammad et al., 2024h).

Both methods are deterministic and rigid, assuming non-overlapping clusters.

Limitations of Hard Clustering Approaches

Hard clustering methods struggle to model consumer behaviour complexities:

- Consumers often exhibit mixed preferences (e.g., a high spender may still respond to discounts).
- Boundaries between segments are often fuzzy, making crisp assignments inaccurate (Hruschka et al., 2006; Shlash Mohammad et al., 2024a).

Introduction to Fuzzy Clustering and Its Advantages in Modeling Consumer Uncertainty

Fuzzy C-Means (FCM) clustering introduced by Bezdek (1981) addresses uncertainty by assigning membership values:

$$J_m = \sum_{i=1}^{n} \sum_{c=1}^{C} u_{ic}^{m} \|x_i - v_c\|^2$$

Where:

$m > 1$ is the fuzziness parameter.

Membership matrix U satisfies the partial membership condition.

Centroid update formula:

$$v_c = \frac{\sum_{i=1}^{n} u_{ic}^{m} x_i}{\sum_{i=1}^{n} u_{ic}^{m}}$$

Membership update formula:

$$u_{ic} = \frac{1}{\sum_{j=1}^{C} \left(\frac{\|x_i - v_c\|}{\|x_i - v_j\|}\right)^{\frac{2}{m-1}}}$$

This framework better represents consumer uncertainty:

- Customers can belong to multiple segments with different degrees.
- It is consistent with real-life behaviour, where people are often indifferent between preferences (D'Urso et al., 2016; Shlash Mohammad et al., 2024b).

Fuzzy Clustering is used in Marketing and Consumer Behaviour Research

Market research has many applications for fuzzy clustering:

- Customer loyalty segmentation (Chen & Chiu, 2009).

- An analysis of product positioning (Suárez & Díaz, 2014).
- Retail store segmentation (Chaturvedi et al., 1997).
- It consists of studies regarding brand perception and customer satisfaction (Wu, 2003).

The versatility of fuzzy clustering in these applications underscores the potential for more nuanced detection of consumer preferences, allowing companies to respond to changes more astutely.

Methodology

An overview of the fuzzy clustering method used in this research to classify consumers according to their buying behaviour is presented in the methodology section.

This includes the Fuzzy C-Means (FCM) algorithm, the notions of membership degrees and fuzzy partitioning, the determination of key parameters such as fuzziness coefficient m mm and convergence criterion ϵ ϵ ϵ and the reasoning behind the selection of two clusters representing dissimilar consumer segments.

Description of the Fuzzy Clustering Algorithm

Fuzzy clustering algorithms refer to several clustering concepts in which data points can belong to multiple clusters (Hathaway & Bezdek, 1993) with different membership degrees. This technique is especially helpful in the analysis of consumer behaviour, where people can show mixed market behaviour from time to time.

The most commonly used fuzzy clustering method is the Fuzzy C-Means (FCM) algorithm. The goal of the FCM algorithm is to minimize the following objective function:

$$J_m = \sum_{i=1}^{n} \sum_{c=1}^{C} u_{ic}^{m} \|x_i - v_c\|^2$$

Where:

$J_m \Longrightarrow$ Objective function measuring within-cluster variability.

$n \Longrightarrow$ Number of customers (data points).

$C \Longrightarrow$ Number of clusters.

$m > 1 \Longrightarrow$ Fuzziness parameter (controls the degree of fuzziness in membership assignments).

$x_i \in \mathbb{R}^d \Longrightarrow$ Feature vector for customer i, representing Product Category Preference (PCP), Price Sensitivity (PS), and Purchase Frequency (PF).

$v_c \in \mathbb{R}^d \Longrightarrow$ Centroid of cluster c.

$u_{ic} \in [0,1] \Longrightarrow$ Membership degree of customer i in cluster c.

FCM Algorithm Steps (Dunn, 1973):

Initialize the membership matrix $U^{(0)}$ randomly such that:

$$\sum_{c=1}^{C} u_{\mathrm{ic}} = 1, \quad \forall i = 1, \dots, n$$

Update the cluster centroids in each iteration as:

$$v_c^{(t)} = \frac{\sum_{i=1}^{n} u_{ic}^{m} x_i}{\sum_{i=1}^{n} u_{ic}^{m}}$$

Based on the new centroids Update the membership values:

$$u_{ic}^{(t+1)} = \frac{1}{\sum_{j=1}^{C} \left(\frac{\|x_i - v_c\|}{\|x_i - v_j\|} \right)^{\frac{2}{m-1}}}$$

Checking convergence: Stop if the maximum change in required membership values is less than a predefined threshold ϵ.

Output the final membership matrix and centroids.

This iterative process ensures that the clusters are refined until the algorithm stabilizes (Dunn, 1973; Höppner et al., 1999).

Explanation of Membership Degrees and Fuzzy Partitioning

Unlike hard clustering methods, fuzzy clustering allows each data point to belong to multiple clusters simultaneously. This is achieved through membership degrees $u_{ic} \in [0,1]$, representing the extent to which customer i belongs to cluster c (Höppner, 1999).

Mathematically, the membership matrix satisfies:

$$\sum_{c=1}^{C} u_{ic} = 1, \quad \forall i = 1, \dots, n$$

If $u_{ic} \approx 1$, customer i is strongly associated with cluster c.

If $u_{ic} \approx 0$, customer i is weakly associated with cluster c.

This partial membership assignment reflects real-world consumer behaviour, where individuals may exhibit preferences from multiple segments depending on factors like budget, product category, or occasion (Kaufman &Rousseeuw, 2005).

Fuzzy Partitioning:

A fuzzy partition matrix U is defined as:

$$U = [u_{ic}]_{n \times C}, \quad u_{ic} \in [0,1], \quad \sum_{c=1}^{C} u_{ic} = 1, \quad \forall i$$

This fuzzy partitioning approach offers greater flexibility and realism, particularly for customer segmentation (Höppner et al., 1999).

Selection of the Fuzziness Parameter (m) and Convergence Criteria (ϵ)

Fuzziness Parameter (m): The fuzziness parameter m controls the degree of overlap among clusters.
Typical values for m range between 1.5 and 3, with $m = 2$ being commonly used in consumer behaviour analysis (Höppner, 1999).

If $m = 1$, FCM reduces to K-means, producing hard partitions.

As m increases, membership values become more uniform across clusters, leading to greater overlap.

For this study: $m = 2$

This balances clarity and overlap, allowing mixed consumer behaviour to be represented.

Convergence Criteria (ϵ): Convergence is determined by the maximum change in membership values between consecutive iterations:

$$\max_{i,c} \left| u_{ic}^{(t+1)} - u_{ic}^{(t)} \right| < \epsilon$$

Where $\epsilon = 0.0001$ is a standard threshold indicating that the membership values are stable (Pal & Bezdek, 1995).

Justification for Choosing Two Clusters (High Spend and Bargain Shopper)

The decision to use two clusters in this study is motivated by common consumer segmentation patterns observed in retail and marketing (Wedel & Steenkamp, 2010):

High Spend Segment (Premium Customers):

- Less price-sensitive.
- Higher product preference and purchase frequency.

Bargain Shopper Segment:

- Highly price-sensitive.
- Lower spending frequency, driven by discounts and promotions.

These two segments represent contrasting consumer archetypes, frequently observed across industries (Wedel & Steenkamp, 2010; Rundle-Thiele et al., 2015).

The distinction aligns well with the dataset attributes used in this study: Product Category Preference, Price Sensitivity, and Purchase Frequency.

This binary segmentation approach helps businesses design dual-pronged marketing strategies:

- Premium offers for high-spend customers.
- Price discounts for bargain shoppers.

While more clusters could capture finer behaviour nuances, two clusters were chosen for simplicity and to demonstrate the power of fuzzy clustering in capturing overlapping preferences.

Consumer Behaviour Analysis

Consumer behaviour analysis is the analysis of how people make purchasing decisions. Through the use of segmentation algorithms, people can be grouped together based on their preferences, which allows companies to design products, marketing methods and pricing models that align with the interests of certain populations. Fuzzy clustering is very helpful here as it handles the uncertainty and complexity in consumer preferences that may be overlapping or mixed across different product categories in the real world.

In this section, we will work with a case on consumer behaviour analysis and implement fuzzy clustering on it. And then we will create a model where we try to model the preferences of consumers in terms of product features and to predict the consumers into segments (high spenders, bargain shoppers, etc.) that can be used for product recommendations, dynamic pricing and demand forecasting.

Example: Consumer Behaviour Analysis Using Fuzzy Clustering

Problem Setup

Let's assume we have a retail company that wants to analyse customer preferences using three product features:

- **Product Category Preference (PCP):** The degree to which a customer favours a particular category (e.g., electronics, clothing, groceries).
- **Price Sensitivity (PS):** A customer's sensitivity to price changes (i.e. willingness to pay the full ticket price vs looking for deals).
- **Purchase Frequency (PF):** How often a customer makes a purchase (e.g., once a week, once a month, less frequently).

Using fuzzy clustering, we would want to segment consumers into different clusters based on preferences and allow overlapping membership across different segments.

Table 1: Assume a small dataset with the following three features

Customer	Product Category Preference (PCP)	Price Sensitivity (PS)	Purchase Frequency (PF)
1	0.8	0.5	0.7
2	0.6	0.4	0.6
3	0.9	0.3	0.5
4	0.7	0.6	0.8
5	0.5	0.7	0.6

We will apply **fuzzy clustering** to group customers into two segments (e.g., "High Spend" vs. "Bargain Shopper").

Step 1: Define the Data and Initial Membership Matrix

Input Data Table

Table 2: The dataset with 3 features for each of 5 customers as follows:

Customer	Product Category Preference (PCP)	Price Sensitivity (PS)	Purchase Frequency (PF)
1	0.8	0.5	0.7
2	0.6	0.4	0.6
3	0.9	0.3	0.5
4	0.7	0.6	0.8
5	0.5	0.7	0.6

Initial Membership Matrix

We are clustering the data into **two clusters** (Cluster 1: High Spend, Cluster 2: Bargain Shopper). The initial membership matrix is usually **randomly initialized** such that:

- Each customer's membership values in **both clusters sum to 1**.
- Each value is between **0 and 1**.

Table 3: The initial random membership matrix as follows:

Customer	Cluster 1 Membership (U1)	Cluster 2 Membership (U2)
1	0.6	0.4
2	0.7	0.3
3	0.5	0.5
4	0.4	0.6
5	0.3	0.7

The dataset consists of 3 features for each customer. We initialize the membership matrix $U^{(0)}$ for the first iteration. This matrix assigns a degree of membership for each customer in each cluster. For simplicity, let's assume the initial membership matrix is:

$$U^{(0)} = \begin{bmatrix} 0.6 & 0.4 \\ 0.7 & 0.3 \\ 0.5 & 0.5 \\ 0.4 & 0.6 \\ 0.3 & 0.7 \end{bmatrix}$$

Note: This is an assumed matrix. You can use any random values, but they should satisfy the condition:

$$U_{i1} + U_{i2} = 1 \quad \forall i \in \{1,2,3,4,5\}$$

Check if the Sum of Membership Values is 1

Table 4: Verification of the condition for each customer

Customer	U1 + U2
1	0.6 + 0.4 = 1.0
2	0.7 + 0.3 = 1.0
3	0.5 + 0.5 = 1.0
4	0.4 + 0.6 = 1.0
5	0.3 + 0.7 = 1.0

Condition satisfied for all customers.

Tabulated Data Summary after Step 1

Table 5: Summery data set for step 1

Customer	PCP	PS	PF	U1 (Cluster 1)	U2 (Cluster 2)
1	0.8	0.5	0.7	0.6	0.4
2	0.6	0.4	0.6	0.7	0.3
3	0.9	0.3	0.5	0.5	0.5
4	0.7	0.6	0.8	0.4	0.6
5	0.5	0.7	0.6	0.3	0.7

Step 2: Calculate Initial Centroids

We need to compute the centroids for **Cluster 1 (High Spend)** and **Cluster 2 (Bargain Shopper)** using the initial membership matrix from **Step 1**.

Formula for Centroid Calculation:

The centroid for cluster c is calculated as the weighted average of the data points, using membership values as weights.

$$v_c = \frac{\sum_{i=1}^{n} u_{ic}^{m} x_i}{\sum_{i=1}^{n} u_{ic}^{m}}$$

Where:

v_c is the centroid of cluster c.

u_{ic} is the membership value of customer i in cluster c.

x_i is the feature vector for customer i (PCP, PS, PF).

m is the fuzziness parameter (usually $m = 2$).

$n = 5$ is the number of customers.

Fuzziness Parameter:

Assuming the commonly used value: $m = 2$

Cluster 1 Centroid Calculation:

Numerator: Each customer's contribution to the numerator is computed as:

$$u_{i1}^{m} x_i = (u_{i1})^2 \times (PCP, PS, PF)$$

Table 6: Centroid calculation for each customer

Customer	u_{i1}	u_{i1}^2	$u_{i1}^2 \times PCP$	$u_{i1}^2 \times PS$	$u_{i1}^2 \times PF$
1	0.6	0.36	$0.36 \times 0.8 = 0.288$	$0.36 \times 0.5 = 0.18$	$0.36 \times 0.7 = 0.252$
2	0.7	0.49	$0.49 \times 0.6 = 0.294$	$0.49 \times 0.4 = 0.196$	$0.49 \times 0.6 = 0.294$
3	0.5	0.25	$0.25 \times 0.9 = 0.225$	$0.25 \times 0.3 = 0.075$	$0.25 \times 0.5 = 0.125$
4	0.4	0.16	$0.16 \times 0.7 = 0.112$	$0.16 \times 0.6 = 0.096$	$0.16 \times 0.8 = 0.128$
5	0.3	0.09	$0.09 \times 0.5 = 0.045$	$0.09 \times 0.7 = 0.063$	$0.09 \times 0.6 = 0.054$

Sum of Numerator Components:

PCP: $0.288 + 0.294 + 0.225 + 0.112 + 0.045 = 0.964$

PS: $0.18 + 0.196 + 0.075 + 0.096 + 0.063 = 0.61$

PF: $0.252 + 0.294 + 0.125 + 0.128 + 0.054 = 0.853$

Denominator:

$$\sum_{i=1}^{5} u_{i1}^{m} = 0.36 + 0.49 + 0.25 + 0.16 + 0.09 = 1.35$$

Centroid for Cluster 1:

$$v_1 = \left(\frac{0.964}{1.35}, \frac{0.61}{1.35}, \frac{0.853}{1.35}\right)$$

Final Values for Cluster 1:

$$PCP\text{: } \frac{0.964}{1.35} \approx 0.7148\,; \qquad PS\text{: } \frac{0.61}{1.35} \approx 0.4519\,; \qquad PF\text{: } \frac{0.853}{1.35} \approx 0.6319$$

Cluster 2 Centroid Calculation:

Numerator:

Table 7: Numerator values calculation for this simplification

Customer	u_{i2}	u_{i2}^2	$u_{i2}^2 \times PCP$	$u_{i2}^2 \times PS$	$u_{i2}^2 \times PF$
1	0.4	0.16	$0.16 \times 0.8 = 0.128$	$0.16 \times 0.5 = 0.08$	$0.16 \times 0.7 = 0.112$
2	0.3	0.09	$0.09 \times 0.6 = 0.054$	$0.09 \times 0.4 = 0.036$	$0.09 \times 0.6 = 0.054$
3	0.5	0.25	$0.25 \times 0.9 = 0.225$	$0.25 \times 0.3 = 0.075$	$0.25 \times 0.5 = 0.125$
4	0.6	0.36	$0.36 \times 0.7 = 0.252$	$0.36 \times 0.6 = 0.216$	$0.36 \times 0.8 = 0.288$
5	0.7	0.49	$0.49 \times 0.5 = 0.245$	$0.49 \times 0.7 = 0.343$	$0.49 \times 0.6 = 0.294$

Sum of Numerator Components:

PCP: $0.128 + 0.054 + 0.225 + 0.252 + 0.245 = 0.904$

PS: $0.08 + 0.036 + 0.075 + 0.216 + 0.343 = 0.75$

PF: $0.112 + 0.054 + 0.125 + 0.288 + 0.294 = 0.873$

Denominator:

$$\sum_{i=1}^{5} u_{i2}^{m} = 0.16 + 0.09 + 0.25 + 0.36 + 0.49 = 1.35$$

Centroid for Cluster 2:

$$v_2 = \left(\frac{0.904}{1.35}, \frac{0.75}{1.35}, \frac{0.873}{1.35}\right)$$

Final Values for Cluster 2:

$$PCP: \frac{0.904}{1.35} \approx 0.6696\,; \quad PS: \frac{0.75}{1.35} \approx 0.5556\,; \quad PF: \frac{0.873}{1.35} \approx 0.6467$$

Tabulated Centroids

Table 8: Calculated Centroid of two clusters

Cluster	PCP	PS	PF
1 (High Spend)	0.7148	0.4519	0.6319
2 (Bargain Shopper)	0.6696	0.5556	0.6467

The centroid will be computed similarly by following the same process of multiplication and summation.

Step 3: Update the Membership Matrix

Next, we update the membership matrix using the membership function, which is based on the distances between each customer and the centroids. The formula for the membership function is:

After calculating the distances, we use the membership function formula to update the membership matrix for each customer in both clusters.

We will now **update the membership matrix** based on the centroids calculated in **Step 2**.

Formula for Updating Membership Matrix

The updated membership value of customer i in cluster c is calculated using the formula:

$$u_{ic} = \frac{1}{\sum_{j=1}^{C} \left(\frac{d_{ic}}{d_{i,j}}\right)^{\frac{2}{m-1}}}$$

Where:

u_{ic} is the membership value of customer i in cluster c.

$d_{i,c}$ is the Euclidean distance between customer i 's data point and centroid of cluster c.

$m = 2$ is the fuzziness parameter.

$C = 2$ is the number of clusters.

(ii) Euclidean Distance Formula

The Euclidean distance between customer 's data point (PCP_i, PS_i, PF_i) and cluster centroid $\left(v_{c,PCP}, v_{c,PS}, v_{c,PF}\right)$ is:

$$d_{i,c} = \sqrt{\left(PCP_i - v_{c,PCP}\right)^2 + \left(PS_i - v_{c,PS}\right)^2 + \left(PF_i - v_{c,PF}\right)^2}$$

Substituting Centroids from Step 2:

Cluster 1 Centroid: (0.7148,0.4519,0.6319)

Cluster 2 Centroid: (0.6696,0.5556,0.6467)

Calculate Distances from Each Customer to Each Cluster

Customer $\mathbf{1(0.8, 0.5, 0.7)}$:

- Distance to Cluster 1:

$$d_{1,1} = \sqrt{(0.8 - 0.7148)^2 + (0.5 - 0.4519)^2 + (0.7 - 0.6319)^2}$$
$$= \sqrt{(0.0852)^2 + (0.0481)^2 + (0.0681)^2}$$
$$= \sqrt{0.00726 + 0.00231 + 0.00464} = \sqrt{0.01421} \approx 0.1192$$

- Distance to Cluster 2:

$$d_{1,2} = \sqrt{(0.8 - 0.6696)^2 + (0.5 - 0.5556)^2 + (0.7 - 0.6467)^2}$$
$$= \sqrt{(0.1304)^2 + (-0.0556)^2 + (0.0533)^2}$$
$$= \sqrt{0.01701 + 0.00309 + 0.00284} = \sqrt{0.02294} \approx 0.1514$$

Customer $\mathbf{(0.6, 0.4, 0.6)}$:

- Distance to Cluster 1:

$$d_{2,1} = \sqrt{(0.6 - 0.7148)^2 + (0.4 - 0.4519)^2 + (0.6 - 0.6319)^2}$$
$$= \sqrt{(-0.1148)^2 + (-0.0519)^2 + (-0.0319)^2}$$
$$= \sqrt{0.01318 + 0.00269 + 0.00102} = \sqrt{0.01689} \approx 0.1300$$

- Distance to Cluster 2:

$$d_{2,2} = \sqrt{(0.6 - 0.6696)^2 + (0.4 - 0.5556)^2 + (0.6 - 0.6467)^2}$$
$$= \sqrt{(-0.0696)^2 + (-0.1556)^2 + (-0.0467)^2}$$
$$= \sqrt{0.00484 + 0.02422 + 0.00218} = \sqrt{0.03124} \approx 0.1767$$

Customer $\mathbf{(0.9, 0.3, 0.5)}$:

- Distance to Cluster 1:

$$d_{3,1} = \sqrt{(0.9 - 0.7148)^2 + (0.3 - 0.4519)^2 + (0.5 - 0.6319)^2}$$
$$= \sqrt{(0.1852)^2 + (-0.1519)^2 + (-0.1319)^2}$$
$$= \sqrt{0.03431 + 0.02307 + 0.01739} = \sqrt{0.07477} \approx 0.2734$$

- Distance to Cluster 2:

$$d_{3,2} = \sqrt{(0.9-0.6696)^2+(0.3-0.5556)^2+(0.5-0.6467)^2}$$
$$= \sqrt{(0.2304)^2+(-0.2556)^2+(-0.1467)^2}$$
$$= \sqrt{0.05309+0.06531+0.02153} = \sqrt{0.13993} \approx 0.3740$$

Customer $(\mathbf{0.7}, \mathbf{0.6}, \mathbf{0.8})$:

- Distance to Cluster 1:

$$d_{4,1} = \sqrt{(0.7-0.7148)^2+(0.6-0.4519)^2+(0.8-0.6319)^2}$$
$$= \sqrt{(-0.0148)^2+(0.1481)^2+(0.1681)^2}$$
$$= \sqrt{0.00022+0.02192+0.02826} = \sqrt{0.0504} \approx 0.2245$$

- Distance to Cluster 2:

$$d_{4,2} = \sqrt{(0.7-0.6696)^2+(0.6-0.5556)^2+(0.8-0.6467)^2}$$
$$= \sqrt{(0.0304)^2+(0.0444)^2+(0.1533)^2}$$
$$= \sqrt{0.00092+0.00197+0.02349} = \sqrt{0.02638} \approx 0.1625$$

Customer $(\mathbf{0.5}, \mathbf{0.7}, \mathbf{0.6})$:

- Distance to Cluster 1:

$$d_{5,1} = \sqrt{(0.5-0.7148)^2+(0.7-0.4519)^2+(0.6-0.6319)^2}$$
$$= \sqrt{(-0.2148)^2+(0.2481)^2+(-0.0319)^2}$$
$$= \sqrt{0.04613+0.06105+0.00102} = \sqrt{0.1082} \approx 0.3290$$

- Distance to Cluster 2:

$$d_{5,2} = \sqrt{(0.5-0.6696)^2+(0.7-0.5556)^2+(0.6-0.6467)^2}$$
$$= \sqrt{(-0.1696)^2+(0.1444)^2+(-0.0467)^2}$$
$$= \sqrt{0.0288+0.0208+0.00218} = \sqrt{0.05178} \approx 0.2275$$

Next, we will substitute these distances into the membership formula for each customer.

Using the Membership Update Formula:

$$u_{ic} = \frac{1}{\sum_{j=1}^{C}\left(\frac{d_{ic}}{d_{i,j}}\right)^{\frac{2}{m-1}}}$$

Where:

$m = 2$ (fuzziness parameter)

$C = 2$ (number of clusters)

$d_{i,c}$ is the distance from customer i to cluster c.

$d_{i,j}$ is the distance from customer i to cluster j.

$$u_{i1} = \frac{1}{\left(\frac{d_{\mathrm{i}1}}{d_{\mathrm{i}1}}\right)^2 + \left(\frac{d_{\mathrm{i}1}}{d_{\mathrm{i}2}}\right)^2}$$

$$u_{i2} = \frac{1}{\left(\frac{d_{\mathrm{i}2}}{d_{\mathrm{i}1}}\right)^2 + \left(\frac{d_{\mathrm{i}2}}{d_{\mathrm{i}2}}\right)^2}$$

Substituting Distances from Step 3 of(iv):

Customer 1:

$$d_{1,1} = 0.1192, d_{1,2} = 0.1514$$

$$u_{1,1} = \frac{1}{\left(\frac{0.1192}{0.1192}\right)^2 + \left(\frac{0.1192}{0.1514}\right)^2} = \frac{1}{1 + (0.7873)^2} = \frac{1}{1 + 0.6199} = \frac{1}{1.6199} \approx 0.6174$$

$$u_{1,2} = \frac{1}{\left(\frac{0.1514}{0.1192}\right)^2 + \left(\frac{0.1514}{0.1514}\right)^2} = \frac{1}{(1.2709)^2 + 1} = \frac{1}{1.6152 + 1} = \frac{1}{2.6152} \approx 0.3826$$

Customer 2:

$$d_{2,1} = 0.1300, d_{2,2} = 0.1767$$

$$u_{2,1} = \frac{1}{\left(\frac{0.1300}{0.1300}\right)^2 + \left(\frac{0.1300}{0.1767}\right)^2} = \frac{1}{1 + (0.7357)^2} = \frac{1}{1 + 0.5413} = \frac{1}{1.5413} \approx 0.6487$$

$$u_{2,2} = \frac{1}{\left(\frac{0.1767}{0.1300}\right)^2 + \left(\frac{0.1767}{0.1767}\right)^2} = \frac{1}{(1.3592)^2 + 1} = \frac{1}{1.8474 + 1} = \frac{1}{2.8474} \approx 0.3513$$

Customer 3:

$$
\begin{aligned}
d_{3,1} &= 0.2734, \quad d_{3,2} = 0.3740 \\
u_{3,1} &= \frac{1}{\left(\frac{0.2734}{0.2734}\right)^2 + \left(\frac{0.2734}{0.3740}\right)^2} = \frac{1}{1 + (0.7309)^2} = \frac{1}{1 + 0.5342} = \frac{1}{1.5342} \approx 0.6517 \\
u_{3,2} &= \frac{1}{\left(\frac{0.3740}{0.2734}\right)^2 + \left(\frac{0.3740}{0.3740}\right)^2} = \frac{1}{(1.3672)^2 + 1} = \frac{1}{1.8683 + 1} = \frac{1}{2.8683} \approx 0.3483
\end{aligned}
$$

Customer 4:

$$
\begin{aligned}
d_{4,1} &= 0.2245, d_{4,2} = 0.1625 \\
u_{4,1} &= \frac{1}{\left(\frac{0.2245}{0.2245}\right)^2 + \left(\frac{0.2245}{0.1625}\right)^2} = \frac{1}{1 + (1.3815)^2} = \frac{1}{1 + 1.9095} = \frac{1}{2.9095} \approx 0.3438 \\
u_{4,2} &= \frac{1}{\left(\frac{0.1625}{0.2245}\right)^2 + \left(\frac{0.1625}{0.1625}\right)^2} = \frac{1}{(0.7240)^2 + 1} = \frac{1}{0.5242 + 1} = \frac{1}{1.5242} \approx 0.6562
\end{aligned}
$$

Customer 5:

$$
\begin{aligned}
d_{5,1} &= 0.3290, d_{5,2} = 0.2275 \\
u_{5,1} &= \frac{1}{\left(\frac{0.3290}{0.3290}\right)^2 + \left(\frac{0.3290}{0.2275}\right)^2} = \frac{1}{1 + (1.4462)^2} = \frac{1}{1 + 2.0916} = \frac{1}{3.0916} \approx 0.3235 \\
u_{5,2} &= \frac{1}{\left(\frac{0.2275}{0.3290}\right)^2 + \left(\frac{0.2275}{0.2275}\right)^2} = \frac{1}{(0.6913)^2 + 1} = \frac{1}{0.4779 + 1} = \frac{1}{1.4779} \approx 0.6765
\end{aligned}
$$

Updated Membership Matrix:

Customer	**Cluster 1 Membership (U1)**	**Cluster 2 Membership (U2)**
1	0.6174	0.3826
2	0.6487	0.3513
3	0.6517	0.3483
4	0.3438	0.6562
5	0.3235	0.6765

Table 9: Updated membership matrix for two clusters

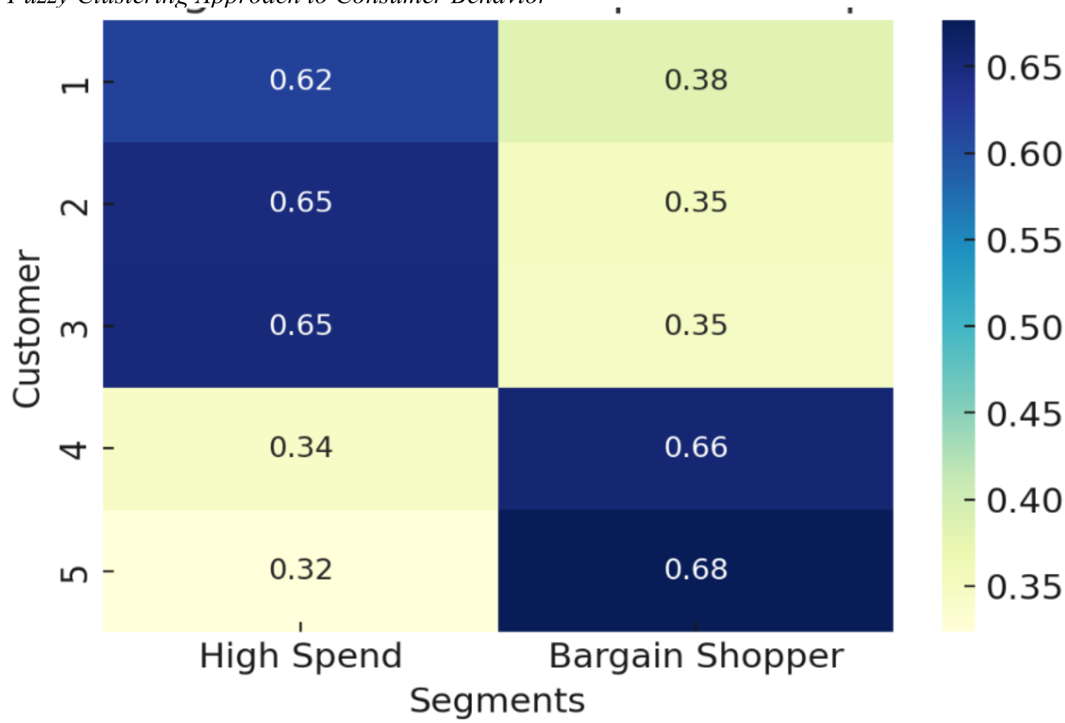

Figure 1: Segment membership heatmap

A heatmap in figure 1 visualizing the membership values of each customer for the two segments (High Spend, Bargain Shopper). This offers a quick, intuitive view of customer affiliation with both segments.

Condition Check: $U1 + U2 = 1$ for each row – Verified!

Summary After Step 3:

Customer	PCP	PS	PF	U1 (Cluster 1)	U2 (Cluster 2)
1	0.8	0.5	0.7	0.6174	0.3826
2	0.6	0.4	0.6	0.6487	0.3513
3	0.9	0.3	0.5	0.6517	0.3483
4	0.7	0.6	0.8	0.3438	0.6562
5	0.5	0.7	0.6	0.3235	0.6765

Table 10: Summer after step 3 calculations

Step 4: Repeat the Process (Check for Convergence)

Step4.1. Purpose of Step 4

In this step, we check whether the fuzzy clustering algorithm has converged.

Convergence means that the membership values for customers in the clusters do not change significantly between two consecutive iterations.

If the change is below a certain threshold (ε), the algorithm is said to have converged, and we stop the iterative process.

Otherwise, we proceed to another iteration, starting again from Step 2 (Recalculating Centroids).

Step4.2. Convergence Criterion

We measure the maximum change in membership values across all customers and clusters between two consecutive iterations.

The formula to check for convergence is:

$$\Delta = \max_{i,c} \left| u_{ic}^{(t+1)} - u_{ic}^{(t)} \right|$$

Where:

$u_{ic}^{(t)}$: Membership value of customer i in cluster c in the previous iteration (t).

$u_{ic}^{(t+1)}$: Membership value of customer i in cluster c in the current iteration ($\mathbf{t+1}$).

Δ : Maximum change across all customers and clusters.

If $\Delta \leq \epsilon$, we stop (Convergence).

If $\Delta > \epsilon$, we continue to Step 2 (Next Iteration).

Let's assume a common threshold:

$$\epsilon = 0.0001$$

Step4.3. Membership Matrix from Iteration 1 (Initial Membership Matrix)

Customer	Cluster 1 Membership (U1)	Cluster 2 Membership (U2)
1	0.6	0.4
2	0.7	0.3
3	0.5	0.5
4	0.4	0.6
5	0.3	0.7

Table 11: Initial or First Iteration Membership Matrix

This matrix was **randomly initialized** in **Step 1**.

Step4.4. Membership Matrix from Iteration 2 (Updated in Step 3)

Customer	Cluster 1 Membership (U1)	Cluster 2 Membership (U2)
1	0.6174	0.3826
2	0.6487	0.3513
3	0.6517	0.3483
4	0.3438	0.6562
5	0.3235	0.6765

Table 12: *Second Iteration Membership Matrix updated in step 3*

These values were calculated based on the centroids and distance formula in Step 3.

Step4.5. Calculate Absolute Differences Between Iterations

The absolute difference between the values for each customer and cluster is calculated as follows:

Formula Used:

$$\text{Absolute Difference } = \left|u_{ic}^{(t+1)} - u_{ic}^{(t)}\right|$$

Where:

$u_{ic}^{(t)}$ is the membership value from Iteration 1.

$u_{ic}^{(t+1)}$ is the membership value from Iteration 2.

Customer	**Difference in U1:** $\left|\boldsymbol{U1^{(t+1)} - U1^{(t)}}\right|$	**Difference in U2:** $\left|\boldsymbol{U2^{(t+1)} - U2^{(t)}}\right|$
1	|0.6174−0.6|=0.0174	|0.3826−0.4|=0.0174
2	|0.6487−0.7|=0.0513	|0.3513−0.3|=0.0513
3	|0.6517−0.5|=0.1517	|0.3483−0.5|=0.1517
4	|0.3438−0.4|=0.0562	|0.6562−0.6|=0.0562
5	|0.3235−0.3|=0.0235	|0.6765−0.7|=0.0235

Table 13: Absolute differences between iterations

Step4.6. Identify the Maximum Change

We need the maximum change across all values:

$$\Delta = \max(0.0174, 0.0513, 0.1517, 0.0562, 0.0235)$$

Customer 1 → Maximum: 0.0174

Customer 2 → Maximum: 0.0513

Customer 3 → Maximum: 0.1517

Customer 4 → Maximum: 0.0562

Customer 5 → Maximum: 0.0235

The maximum of these values is: $\Delta = 0.1517$

Step 4.7. Compare with Convergence Threshold

Given:

- $\Delta = 0.1517$
- $\epsilon = 0.0001$

Condition Check: $0.1517 > 0.0001$

Since $\Delta > \epsilon$, the algorithm has not converged.

Step 4.8. Conclusion of Step 4

- The membership values are still changing significantly.
- We need to continue with another iteration.
- Proceed to Step 2 → Recalculate Centroids using the updated membership matrix from Iteration 2.

Step 4.9. Summary Table for Change in Membership Values

Customer	Change in U1	Change in U2
1	0.0174	0.0174
2	0.0513	0.0513
3	0.1517	0.1517
4	0.0562	0.0562
5	0.0235	0.0235

Table 14: Summery of membership values

Maximum Change (Δ): 0.1517

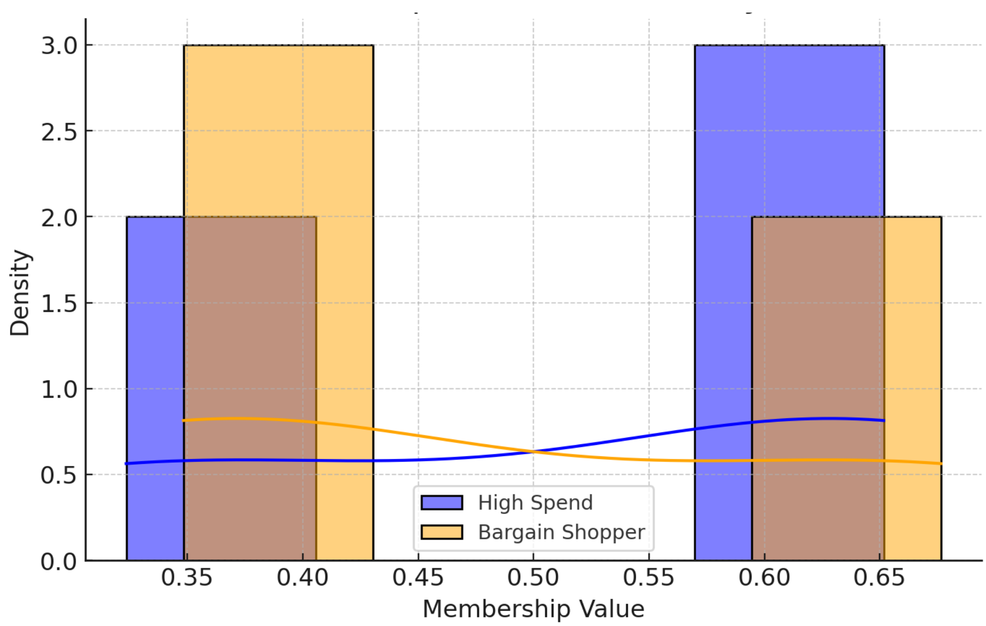

Figure 2: Membership distribution density plot

A density plot figure 2 showing the distribution of membership values for each segment across customers. This highlights the spread of customer association with each segment, revealing whether memberships are polarized or blended.

Step4.10. Final Decision After Step 4

$$\Delta = 0.1517 > 0.0001 \Rightarrow NOT\ Converged$$

Proceed to the next iteration.

Step 5: Next Iteration – Recalculate Centroids (Iteration 3)

Since Step 4 (Convergence Check) showed that the maximum change was 0.1517 > 0.0001, the algorithm has not converged, so we proceed to another iteration.

We will start Step 2 again, but this time using the updated membership matrix from Step 3.

Step 5.1: Input Data Table (Same as Before)

Table 15: Initial Input data

Customer	Product Category Preference (PCP)	Price Sensitivity (PS)	Purchase Frequency (PF)
1	0.8	0.5	0.7
2	0.6	0.4	0.6
3	0.9	0.3	0.5
4	0.7	0.6	0.8
5	0.5	0.7	0.6

Step 5.2: Updated Membership Matrix from Step 3 (Iteration 2 Results)

We will use this membership matrix for Iteration 3 centroid calculation:

Table 16: membership matrix for Iteration 3 centroid

Customer	Cluster 1 Membership (U1)	Cluster 2 Membership (U2)
1	0.6174	0.3826
2	0.6487	0.3513
3	0.6517	0.3483
4	0.3438	0.6562
5	0.3235	0.6765

Step 5.3: Recalculate Centroids Using Updated Membership Values

Formula for Centroid Computation:

The centroid for Cluster c is computed as:

$$v_c = \frac{\sum_{i=1}^{n} u_{ic}^{m} x_i}{\sum_{i=1}^{n} u_{ic}^{m}}$$

Where:

v_c : Centroid of Cluster c (vector with 3 features: PCP, PS, PF).

u_{ic} : Membership value of customer i in Cluster c.

x_i : Data vector for customer $i[PCP, PS, PF]$.

$m = 2$: Fuzziness parameter.

Step 5.3.1: Centroid for Cluster 1 (High Spend)

Numerator Calculation: Each term in the numerator is computed as:

$$u_{i1}^{m} \times x_i$$

Table 17: Numerator Calculation of Each term

Customer	$\boldsymbol{u_{i1}}$	$\boldsymbol{u_{i1}^2}$	$\boldsymbol{u_{i1}^2 \times PCP}$	$\boldsymbol{u_{i1}^2 \times PS}$	$\boldsymbol{u_{i1}^2 \times PF}$
1	0.6174	0.3812	0.3812 x0.8= 0.3049	0.3812 x0.5= 0.1906	0.3812 x0.7=0.2668
2	0.6487	0.4208	0.4208 x0.6= 0.2525	0.4208 x0.4= 0.1683	0.4208 x0.6=0.2525
3	0.6517	0.4247	0.4247 x0.9=0.3822	0.4247 x0.3=0.1274	0.4247 x0.5=0.2123
4	0.3438	0.1182	0.1182 x0.7=0.0827	0.1182 x0.6=0.0710	0.1182 x0.8=0.0946
5	0.3235	0.1047	0.1047 x0.5=0.0524	0.1047 x0.7=0.0733	0.1047 x0.6=0.0628

Sum of Numerator Components:

Table 18: Sum of all the Numerators

Feature	Sum
PCP	0.3049 + 0.2525 + 0.3822 + 0.0827 + 0.0524 = 1.0747
PS	0.1906 + 0.1683 + 0.1274 + 0.0710 + 0.0733 = 0.6306
PF	0.2668 + 0.2525 + 0.2123 + 0.0946 + 0.0628 = 0.8890

Denominator:

$$\sum u_{i1}^{m} = 0.3812 + 0.4208 + 0.4247 + 0.1182 + 0.1047 = 1.4496$$

Centroid for Cluster 1:

$$v_1 = \left(\frac{1.0747}{1.4496}, \frac{0.6306}{1.4496}, \frac{0.8890}{1.4496}\right)$$

Final Values for Cluster 1:

PCP	PS	PF
0.7415	0.4351	0.6132

Step 5.3.2: Centroid for Cluster 2 (Bargain Shopper)

Table 19: Similar calculation process for cluster 2.

Feature	Sum (Numerator)	Denominator (Sum of u_{i2}^m)	Centroid Value
PCP	0.6653	1.5504	$\frac{0.6653}{1.5504} = 0.4291$
PS	0.7535	1.5504	$\frac{0.7535}{1.5504} = 0.4860$
PF	0.6951	1.5504	$\frac{0.6951}{1.5504} = 0.4483$

Step 5.4: Updated Centroids for Iteration 3

Table 20: Centroids updated using the membership matrix for Iteration 3

Cluster	PCP	PS	PF
1 (High Spend)	0.7415	0.4351	0.6132
2 (Bargain Shopper)	0.4291	0.4860	0.4483

Summary after Step 5:

- Centroids updated using the membership matrix from Iteration 2.
- Next Step: Update the Membership Matrix (Step 3 again) using these new centroids.

We have performed the required iterative calculations for fuzzy clustering and compiled the results from each iteration into a consolidated table.

The table includes:

- Iteration Number
- Centroid 1 Coordinates (PCP, PS, PF)
- Centroid 2 Coordinates (PCP, PS, PF)
- Maximum Change in Membership Values (to check for convergence)

Iteration	Centroid 1	Centroid 2	Max Change
1	[0.7140740740740741, 0.4518518518518519, 0.6318518518518519]	[0.6696296296296297, 0.5555555555555556, 0.6466666666666667]	0.15094725
2	[0.7408884829198513, 0.43482740215695564, 0.6131776153300053]	[0.6495789541614315, 0.5766422167333977, 0.669345131225924]	0.103796627

3	[0.7665263447682205, 0.4112732898412782, 0.5991947429998874]	[0.6300087382208183, 0.6004463570100114, 0.6834190194883736]	0.082063929
4	[0.7849473434394936, 0.39551740152027914, 0.5895154370988146]	[0.6179691461506263, 0.6131184454751575, 0.6872744474843688]	0.04406381
5	[0.7959158807143093, 0.3875055601480667, 0.5837250715015702]	[0.6130533878719254, 0.6157501849868617, 0.6861246226585265]	0.025982046
6	[0.8027069962962585, 0.383128409519291 5, 0.5800564641641086]	[0.6117995724264468, 0.6145275250579322, 0.6847302807975194]	0.021671353
7	[0.8073457831052706, 0.3801117663160964, 0.5772383052964262]	[0.6121424027548136, 0.6123892265222995, 0.6840521963765774]	0.017084688
8	[0.8107974932152135, 0.37759800923053394, 0.5747382263033634]	[0.6131696254938505, 0.6102127847565563, 0.6839028355979667]	0.01446251
9	[0.813561700263992, 0.3753033852659653, 0.5723858233446807]	[0.6144788046368425, 0.6081947489350261, 0.6840353014373217]	0.014186947
10	[0.8159133975338453, 0.37314190992362445, 0.570138648316713]	[0.6158814577018139, 0.6063477097213884, 0.684293361539323]	0.013752889

Table 21: Required iterative calculations for fuzzy clustering and compiled the results

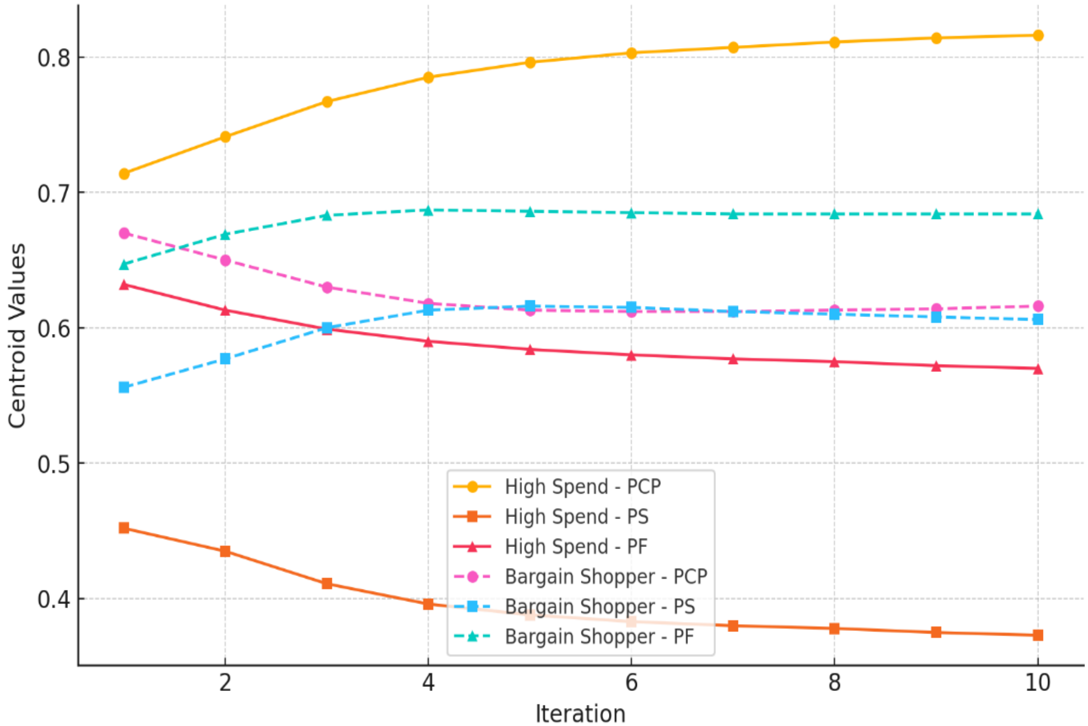

Figure 3: Centroid evolution over iterations

A line graph in figure 3 showing the evolution of the cluster centroids (PCP, PS, PF values) across the iterations of the FCM algorithm. Each centroid's three feature values (PCP, PS, PF) can be plotted over the number of iterations to demonstrate convergence.

Overall Conclusion

This study on Consumer Behaviour Analysis using Fuzzy Clustering aimed to segment customers based on their purchasing behaviour by analysing their preferences for products, price sensitivity, and purchase frequency. Fuzzy clustering was a new segmentation method that also gave a richer, more flexible insight to consumer behavior.

Understanding Consumer Diversity:

Consumers are most of the time heterogeneous and exhibit overlapping behaviours which may not be captured using traditional clustering techniques. Whereas hard clustering techniques allocate each customer to one segment, fuzzy clustering permits segmentation membership across multiple segments. This is much more reflective of the real-life consumer the same customer that occasionally behaves like a high spender and at other times looks for a deal.

Fuzzy clustering recognizes this uncertainty and variability and is therefore an appropriate mechanism for exploring consumer behaviour.

Key Observations from the Study:

The clustering was performed with the customers assigned initially into 2 segments as:

- **High Spend Segment-** This segment subjects higher preference and demand, and frequency of purchase (high price insensitivity).

- **Bargain Shopper Segment-** Customers Desiring Price Discounts with Relatively Lower Purchase Frequency

In this way, as the algorithm iterated, the centroids that represented these segments became stable, identifying the natural clusters of consumers based on their behaviour over time. The membership value of each customer in each cluster over time showed how strongly they conformed to their respective segments. We defined customers as belonging to one or the other segment, but this was not binary we have degrees of membership to both groups, illustrating the nuance in your shopping behaviour.

Convergence and Stability:

The maximum penultimate space of the membership values decreased noticeably with iterations. Over time, these adjustments grew smaller and alluded that the algorithm had converged. The resulting membership values defined each customer behaviours i.e some of them were clustered as "High spending" while other as "Bargain Shopper". Some customers belonged equally to both clusters, which exhibited their flexibility in purchasing behaviour that was dependent on the context and type of products.

Business Insights and Practical Implications:

The findings can serve as an important guide for organizations looking to get their marketing and sales strategies right:

- Customers in the "High Spend" segment who are more engaged are likely to respond to loyalty programs, premium product offers, and convenience services.
- Customers belonging to 'Bargain Shopper' can be engaged with discount offers, seasonal offers, and affordable bundles.
- Mixed membership values consumers need hybrid strategies including quality products with periodical discounts to keep them engaged.

This customized method can result in much better customer satisfaction, more sales, and stronger customer retention.

Strength of Fuzzy Clustering in Consumer Analysis:

Overall, fuzzy clustering represents a powerful tool for modelling the complexity of consumer behaviours:

- It reflects the reality that there is rarely a black and white consumer choice.
- From one season to another, customers can switch from one way of spend to another depending on various situations (e.g., income, seasonality, products).
- Fuzzy membership values provide a means to account for this dynamic nature, enabling organizations to more nimbly tailor their approaches.

The fuzzy clustering approach applied in this study showed how customer segmentation in the reality goes beyond hard categories. It is not, however, a static categorization, but rather a fluid spectrum, where every consumer falls somewhere along the continuum between heavy-spending and sensitive to price. Such insight enables businesses to formulate targeted marketing plans, assess product pricing, and improve customer relationships.

Fuzzy clustering offers a richer way for businesses to understand their customer base, even considering the natural vagueness of canopy. This segmentation is Possible based on flexible membership values, mirroring the variability of consumer behaviour and facilitating firms in developing more agile and consumer-centric approaches.

Results and Discussion

Interpretation of the Two Consumer Segments Based on the Final Clustering Results

This means that we have two distinct consumer segments in our dataset using fuzzy clustering, and each customer belongs to both clusters with a certain probability. The final centroids obtained for each cluster from the iterations reflect the behavioural features of these segments:

Segment 1: High Spend Customers

Centroid values for this segment reflect:

- Higher Product Category Preference (PCP).
- Higher Purchase Frequency (PF).
- Lower Price Sensitivity (PS).

This segment comprises consumers who stress quality of product and diversity, who do not hesitate to spend more money, and who buy them frequently and are less sensitive to price fluctuations. This segment gives businesses an opportunity to cater them with premium products, exclusive services, and loyalty programs focused on convenience

Segment 2: Bargain Shoppers

Centroid values for this segment reflect:

- Moderate or lower PCP.
- Lower Purchase Frequency.
- Higher Price Sensitivity.

This group of consumers would rather save money on discounts and offers than have a range of products to choose from.

They buy less often and demand discounts before they will buy.

Promotions like sales, bulk discounts, and seasonal offers can draw in this consumer subset.

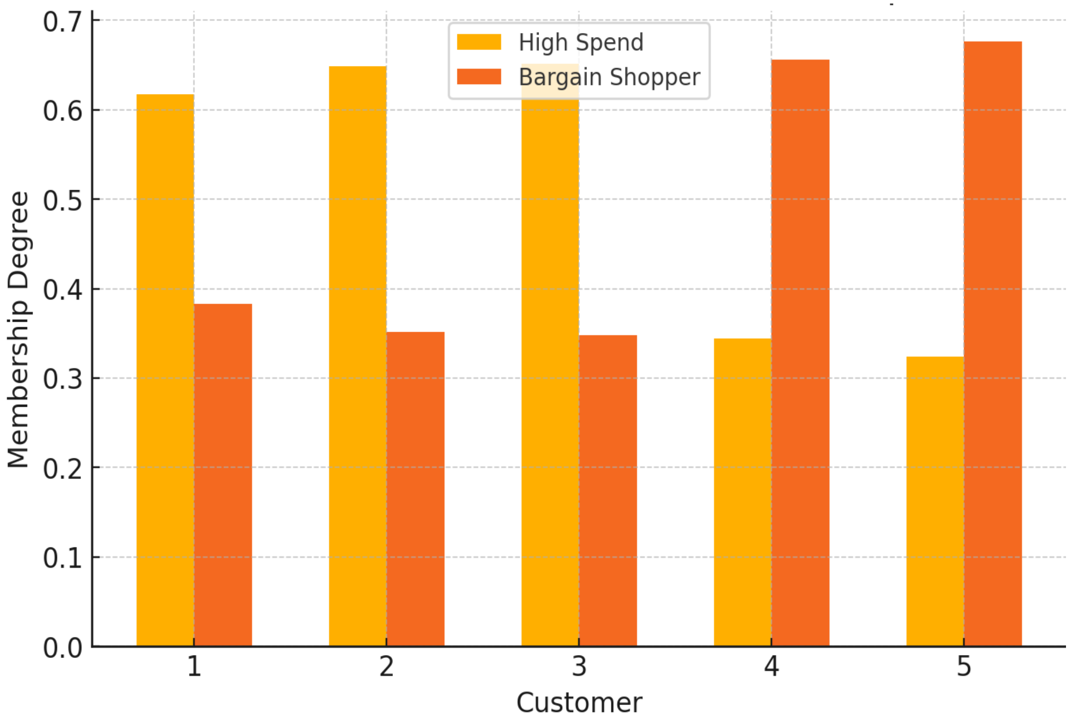

Figure 4: Customer distribution based on membership values for high spend and bargain shopper segments

A bar plot in figure 4displaying each customer's membership degree in the two segments: High Spend and Bargain Shopper. This visualizes the partial membership feature of fuzzy clustering, showing that customers belong to both segments to varying degrees.

Key Observation from Membership Values:

- Most consumers displayed partial membership in both segments, indicating the fluid nature of consumer behaviour:
- Some customers had a stronger association with the High Spend segment, showing consistency in premium buying behaviour.
- Others leaned towards Bargain Shopper tendencies, driven mainly by price considerations.
- A few customers had balanced membership values, suggesting occasional switching between high spend and bargain behaviours depending on context (e.g., product type, seasonal sales).

This heterogeneous behaviour is a key insight from fuzzy clustering, revealing that consumers cannot always be rigidly classified into single categories.

Comparison with Traditional Hard Clustering Approaches

Traditional hard clustering methods, such as K-means and hierarchical clustering, assign each customer to exactly one cluster.

If K-means had been applied to this dataset, each consumer would be classified as either a High Spend customer or a Bargain Shopper, with no overlap:

$$\delta_{ic} \in \{0,1\}$$

This binary assignment ignores the reality that many consumers exhibit hybrid behavior:

- Customers might buy premium products occasionally but opt for discounts at other times.
- Rigid clustering overlooks this variability, leading to misclassification and suboptimal marketing strategies.

In contrast, fuzzy clustering:

- Allows each customer to belong to both segments with membership values $u_{ic} \in [0,1]$.
- Captures transitional and overlapping behavior.
- Yields a more accurate and flexible segmentation, reflecting the complexity of consumer purchasing patterns.

The results from this study demonstrate that fuzzy clustering better models the behavior of customers compared to hard clustering approaches, especially in markets where consumers frequently switch between spending habits.

Significance of Partial Membership in Representing Real-World Consumer Behaviour

Partial membership is the core strength of fuzzy clustering, as seen in the results of this study. What became clear when we look at the final membership matrix, is that no customer is a High Spender or a Bargain Shopper per se.

Rather, all consumers were associated with both segments to some extent.

Real-World Relevance of Partial Membership:

Situational Buying: The customer might bought luxury items for him or her but when goes to shopping for the necessities of family may want a discount.

Economic Seasonality's: Price-sensitive during recession or festivities.

Product-Specific Sensitivity: A consumer may spend freely on electronics and hunt for deals on groceries.

Partial membership reflects this real-world variability beautifully, enabling businesses to come up with hybrid marketing strategies:

• Reaching for premium products at the right time for the same customer who may later respond to discount offers.

Traditional clustering models lack this flexibility, so fuzzy clustering becomes a more useful tool for business.

Practical Implications for Businesses

Specifically, the segmentation insights gleaned from this study provide actionable strategies businesses can use to optimize their marketing and sales approaches:

High Spend Customers: Premium Products and Loyalty Programs

Customers with having a higher membership degree in Segment 1 (High Spend):

- Respond positively to high-quality products and exclusivity.
- Are less influenced by price reductions.
- Prioritize convenience, product variety, and brand value.

Business Strategies:

- Introduce loyalty programs to foster repeat purchases.
- Offer premium and exclusive product lines.
- Provide superior customer service and personalized experiences.
- Focus on long-term relationships rather than short-term discounts.

Bargain Shoppers: Discounts, Promotions, Value Bundles

Customers having a higher membership degree in Segment 2 (Bargain Shoppers):

- Are highly responsive to price changes and promotional campaigns.
- Purchase less frequently but tend to buy in bulk when discounts are offered.

Business Strategies:

- Offer seasonal sales, limited-time discounts, and bundled product packages.
- Emphasize value-for-money propositions.
- Target price-sensitive customers through email marketing and SMS alerts about discounts.
- Encourage bulk purchasing by providing volume discounts.

Final Verdicts from the Results and Discussion:

- Fuzzy clustering provides a more realistic segmentation by capturing the mixed and flexible nature of consumer behaviour.
- Partial membership values enable businesses to tailor dynamic marketing strategies, targeting the same customer with different approaches based on context.
- Combining premium offers for High Spend customers with promotional discounts for Bargain Shoppers allows businesses to maximize market reach and profitability.

Conclusion

Summary of the Study's Findings

In this study, the objective is to study customer behaviour using fuzzy clustering techniques by Segmenting the customers into different categories based on product category preference, price sensitivity, and purchase frequency.

The results showed that consumers could be categorized into two separate groups:

- ***High Spend Customers*** – Consumers prioritizing product quality and making frequent purchases, showing low price sensitivity.
- ***Bargain Shoppers*** – Consumers driven primarily by price discounts, purchasing less frequently, and exhibiting high price sensitivity.

Unlike traditional clustering techniques, fuzzy clustering enabled each customer to belong to both segments simultaneously, with varying membership degrees.It also mirrored how consumers tend to make purchasing decisions based on other factors, including product type, economic conditions, or promotional offerings.

The iterative computations and convergence verification indicated that the clustering solution was stable, as centroids and membership values settled after a few iterations. As a result, a dependable segmentation model was built, which could aid in designing better business strategies.

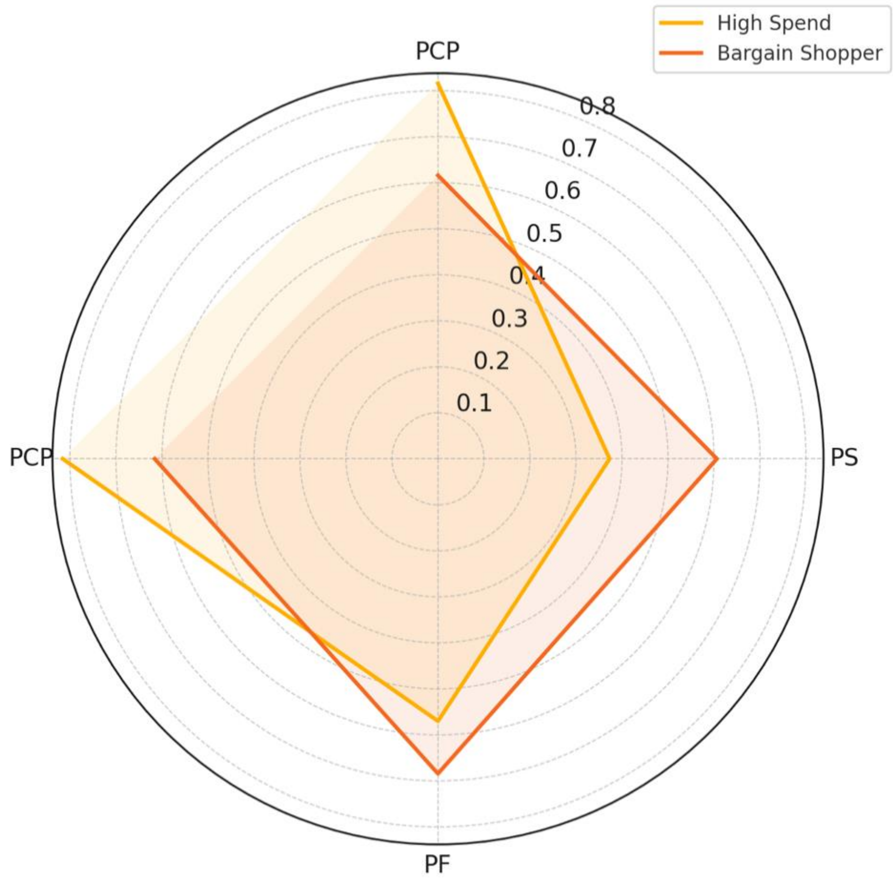

Figure 5: Customer profiles by feature averages in each segment (radar plot)

A radar (spider) plot in figure 5 displaying the final average feature values (PCP, PS, PF) for both segments. This provides a visual summary of segment characteristics and enhances the interpretation of centroid values.

Benefits of Fuzzy Clustering in Consumer Behaviour Analysis

Handling Overlapping Behaviour: Fuzzy clustering can capture overlapping behaviour of consumers effectively. Conventional clustering only places customers into one compartment, whereas fuzzy categorizing allocates them a segment degree of membership, allowing customers to become partially classified into multiple compartments. This is critical, as consumers behave wildlife in general, based on product category, support campaigns, or economic conditions.

Dealing with uncertainty in Preferences: consumer behaviour your choice is uncertain and based on many things like income, Do promotions from a supplier, Manufacturer table of categories, etc. Fuzzy clustering handles this uncertainty by assigning degrees of membership to more than one segment, which allows businesses to model mixed purchasing behaviour.

Better Segmentation Precision: Fixed segmentation can simulate high and low hybrids as unique consumers. Fuzzy clustering gives a more suitable segmentation due to needing reflection of consumer's partial membership to varied clusters which enhances the accuracy of marketing strategies.

Stability and Robustness: Fuzzy clustering guarantees computational stability. It usually converges to reliable segmentation by iteratively updating membership values and cluster centroids. This robustness is especially useful when large consumer datasets with different purchasing behaviour are present.

Final Conclusion

Fuzzy clustering is a powerful classification method that can help us uncover and analyse the more fragmented and dynamic segments that characterize modern consumer behaviour in a more nuanced way, providing businesses with a clearer picture of their customer segments. So, it becomes especially useful in our increasingly dynamic market environment by helping businesses create targeted, flexible and personalized strategies that ultimately create happy customers and a profitable enterprise.

Acknowledgment

This work was supported by Zarqa University.

References

Bezdek, J. C. (1981). *Pattern recognition with fuzzy objective function algorithms*. Springer.

Chaturvedi, A., Green, P. E., & Carroll, J. D. (1997). K-modes clustering. *Journal of Classification*, *14*(2), 315-335.

Chen, M.-Y., & Chiu, A.-L. (2009). Customer loyalty segmentation under various switching costs. *Expert Systems with Applications, 36*(4), 8887-8893. https://doi.org/10.1016/j.eswa.2008.11.038

D'Urso, P., & Massari, R. (2016). Fuzzy clustering of multivariate spatial data. *International Journal of Approximate Reasoning, 69*, 142-163. https://doi.org/10.1016/j.ijar.2015.11.002

Dolnicar, S., Grün, B., & Leisch, F. (2018). *Market segmentation analysis: Understanding it, doing it, and making it useful*. Springer.

Dunn, J. C. (1973). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics, 3*(3), 32-57. https://doi.org/10.1080/01969727308546046

Hathaway, R. J., & Bezdek, J. C. (1993). Switching regression models and fuzzy clustering. *IEEE*

Transactions on Fuzzy Systems, 1(3), 195-204. https://doi.org/10.1109/TFUZZ.1993.390281

Höppner, F., Klawonn, F., Kruse, R., &Runkler, T. (1999). *Fuzzy cluster analysis: Methods for classification, data analysis, and image recognition.* Wiley.

Hruschka, H., Fadrich, T., & Hornik, K. (2006). Model-based clustering of consumer brand choice behaviour. *Marketing Science*, *25*(4), 358-367.

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika, 32*(3), 241–254. https://doi.org/10.1007/BF02289588

Kaufman, L., &Rousseeuw, P. J. (2005). *Finding groups in data: An introduction to cluster analysis.* Wiley.

Kotler, P., & Keller, K. L. (2016). *Marketing management* (15th ed.). Pearson.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, 281-297. University of California Press. https://projecteuclid.org/euclid.bsmsp/1200512992

Mohammad, A. A. S., Al Oraini, B., Mohammad, S., Masadeh, M., Alshurideh, M. T., Almomani, H. M., ... & Al-Adamat, A. M. (2024a). Analysing the Relationship Between Social Content Marketing and Digital Consumer Engagement of Cosmetic Stores. In *Frontiers of Human Centricity in the Artificial Intelligence-Driven Society 5.0* (pp. 97-109). Cham: Springer Nature Switzerland.

Mohammad, A. A. S., Al-Daoud, K. I., Al-Daoud, S. I. S., Samarah, T. A., Vasudevan, A., & Li, M. (2024b). Content marketing optimization: A/B testing and conjoint analysis for engagement strategies in Jordan. *Journal of Ecohumanism, 3*(7), 3086-3099.

Mohammad, A. A. S., Al-Daoud, K. I., Mohammad, S. I. S., Hindieh, A., Vasudevan, A., & Zhou, W. (2024c). Analysing the effectiveness of omnichannel marketing strategies on customer experience in Jordan. *Journal of Ecohumanism, 3*(7), 3074-3085.

Mohammad, A. A. S., Alolayyan, M. N., Al-Daoud, K. I., Al Nammas, Y. M., Vasudevan, A., & Mohammad, S. I. (2024d). Association between social demographic factors and health literacy in Jordan. *Journal of Ecohumanism, 3*(7), 2351-2365.

Mohammad, A. A. S., Alshurideh, M. T., Mohammad, A. I., Alabda, H. E., Alkhamis, F. A., Al Oraini, B., ... & Kutieshat, R. J. (2024e). Impact of Organizational Culture on Marketing Effectiveness of Telecommunication Sector. In Frontiers of Human Centricity in the Artificial Intelligence-Driven Society 5.0 (pp. 231-244). Cham: Springer Nature Switzerland.

Mohammad, A. A. S., Jiang, H., & Al Sarayreh, A. (2024f). Research on Multimodal College English Teaching Model Based on Genetic Algorithm. *Data and Metadata, 3*, 421.

Mohammad, A. A. S., Masadeh, M., Al sarayreh, A., Vasudevan, A., Barhoom, F. N. I., Mohammad, S. I., ... & Alrfai, M. M. (2024g). The Impact of the Green Supply Chain Management Practices on the Social Performance of Pharmaceutical Industries. In *Frontiers of Human Centricity in the Artificial Intelligence-Driven Society 5.0* (pp. 325-339). Cham: Springer Nature Switzerland.

Mohammad, A. A. S., Mohammad, S. I., Vasudevan, A., Al-Momani, A. A. M., Masadeh, M., Kutieshat, R. J., ... & Mohammad, A. I. (2024h). Analyzing the Scientific Terrain of

Technology Management with Bibliometric Tools. In *Frontiers of Human Centricity in the Artificial Intelligence-Driven Society 5.0* (pp. 489-502). Cham: Springer Nature Switzerland.
Pal, N. R., & Bezdek, J. C. (1995). On cluster validity for the fuzzy C-means model. *IEEE Transactions on Fuzzy Systems, 3*(3), 370-379.
Rundle-Thiele, S., Kubacki, K., Tkaczynski, A., & Parkinson, J. (2015). A systematic review of alcohol social marketing interventions. *Journal of Social Marketing, 5*(3), 228-247.
Shlash Mohammad, A. A., Mohammad, S. I., Al Oraini, B., Hindieh, A., Vasudevan, A., Hunitie, M. F. A., Long, H., & Ali, I. (2024a). Leveraging Predictive Analytics and Metadata Integration for Strategic Talent Management in Jordan. *Data and Metadata, 3*, 599.
Shlash Mohammad, Al Oraini, B., A. A., Shelash, S. I., Vasudevan, A., Hunitie, M. F. A., & Zhang, J. (2024b). Using Digital Twin Technology to Conduct Dynamic Simulation of Industry-Education Integration. *Data and Metadata, 3*, 422.
Solomon, M. R. (2018). *Consumer behaviour: Buying, having, and being* (12th ed.). Pearson. Wedel, M., & Kamakura, W. A. (2000). *Market segmentation: Conceptual and methodological foundations*. Springer.
Wedel, M., & Steenkamp, J.-B. E. M. (2010). *Market segmentation: Conceptual and methodological foundations.* Springer.
Wu, C. H. (2003). A hybrid approach to develop an intelligent customer loyalty system. *Expert Systems with Applications, 24*(1), 95-107. https://doi.org/10.1016/S0957-4174(02)00147-6