



Article

Analyzing the Dynamics of Customer Behavior: A New Perspective on Personalized Marketing through Counterfactual Analysis

Mona Ebadi Jalal * and Adel Elmaghraby

Department of Computer Science and Engineering, University of Louisville, Louisville, KY 40292, USA;
adel@louisville.edu

* Correspondence: mona.ebadijalal@louisville.edu

Abstract: The existing body of research on dynamic customer segmentation has primarily focused on segment-level customer purchasing behavior (CPB) analysis to tailor marketing strategies for distinct customer groups. However, these approaches often lack the granularity required for personalized marketing at the individual level. Moreover, the analysis of customer transitions between different groups has largely been overlooked. This study addresses these gaps by developing an efficient framework that enables businesses to forecast customer behavior, assess the impact of various strategies on each customer separately, and analyze customer transition between segments. This can facilitate providing personalized marketing strategies, fostering a gradual transition toward a desired customer status, and enhancing the overall marketing precision. In this study, we employ time series feature vectors encompassing recency, frequency, monetary value, and lifespan, applying the K-means algorithm with a range of distance metrics for customer segmentation along with classification algorithms to predict customer behavior. Leveraging counterfactual analysis, we establish a solution for analyzing customer transitions between groups and evaluating personalized marketing strategies. Our findings underscore the superior performance of the Euclidean distance metric, closely followed by the Manhattan distance, in distinguishing the patterns in time series customer behavior, with logistic regression excelling in predicting customer status. This study enables decision-makers to forecast the impact of diverse marketing strategies on customer behavior which facilitates customer retention and engagement through well-informed decisions.

Keywords: customer segmentation; customer behavior; RFM model; time series clustering; counterfactual analysis; personalized marketing; machine learning



Citation: Ebadi Jalal, M.; Elmaghraby, A. Analyzing the Dynamics of Customer Behavior: A New Perspective on Personalized Marketing through Counterfactual Analysis. *J. Theor. Appl. Electron. Commer. Res.* **2024**, *19*, 1660–1681. <https://doi.org/10.3390/jtaer19030081>

Academic Editor: Carlos Orús

Received: 6 May 2024

Revised: 12 June 2024

Accepted: 18 June 2024

Published: 27 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Customers include both individuals and businesses that purchase products or services from a company. Maintaining a good relationship with customers over a long period can lead to increased profits from existing customers for an enterprise [1]. According to Kotler and Armstrong [2], while attracting customers is undoubtedly important, the retention of customers holds even greater significance. This is due to the fact that losing a customer not only results in an immediate loss but also forfeits the potential lifetime value of their purchases. To enhance customer retention and boost profitability, each organization must tailor marketing strategies, efficiently allocate resources, and effectively meet the different needs of its customer base [3,4]. In recent decades, with the rise of personalized marketing in e-commerce, traditional mass marketing is becoming increasingly obsolete [5]. Personalized marketing should utilize individual-level information to tailor interactions, enhancing customer experience and marketing effectiveness for a competitive advantage in a knowledge-driven world [6]. The shift towards personalized marketing strategies requires a profound understanding of each customer's unique behavioral characteristics. Segmentation is a commonly employed method to achieve this objective [5,7].

Customers often exhibit diverse preferences, which makes customer segmentation a valuable strategy for effectively managing companies' relationships with their customers [8,9]. Customer segmentation is a popular tool that involves grouping customers with similar characteristics and attributes from larger, more heterogeneous groups of customers [9–11]. Traditional customer segmentation methods, such as those described by Calvet et al. [12], typically categorize customers based on descriptive variables like demographic attributes. However, when demographic data are unavailable or cannot be inferred from the existing data, these conventional segmentation techniques become impractical [13]. The recency, frequency, and monetary (RFM) model is a widely used value-based approach to conducting behavioral customer segmentation [5,14–21]. Specifically, “recency” pertains to the time elapsed since the last purchase, “frequency” signifies the number of purchases within a defined time frame, and “monetary” reflects the amount of money spent during this specified period [22]. These three variables fall under the category of behavioral attributes and can serve as segmentation criteria by evaluating customer attitudes towards the product, brand, benefits, or even their loyalty, all gleaned from the database [23].

The main challenge in the current literature on customer segmentation, including studies that utilize the RFM model, lies in typically employing a single time frame to encompass customer behavior, often referred to as “static segmentation” (as seen in examples such as [24–29]). The primary limitation of static segmentation approaches lies in their inability to model the dynamic behavior of customers and uncover significant trends and patterns [30–32]. A few studies adopt a dynamic approach to customer segmentation [31]. These approaches entail either monitoring customer behavior trajectories over time using techniques like sequential rule mining [32] or representing customer behavior as time series and subsequently applying time series analytical methods [15,19]. The main advantage of dynamic customer segmentation utilizing time series features over other approaches lies in its predictive capability. The primary limitation associated with sequential rule mining approaches lies in their descriptive nature, as they are designed to elucidate historical trends in customer behavior rather than being suitable for forecasting future customer behavior [17]. In summary, the main goal of earlier studies on customer segmentation utilizing static approaches was to segment customer behavior, analyze each segment, and identify prevailing trends. In dynamic approaches, researchers strive not only to achieve these objectives but also to predict future customer behavioral patterns. This information can then be utilized by companies' marketing departments to make well-informed decisions tailored to each customer group. Table 1 summarizes research studies on dynamic customer segmentation using time series features.

Table 1. Recent studies on dynamic customer segmentation utilizing time series features.

Study	Features	Methodology	Dataset	Conclusion
Song et al. [14] 2017	Recency, frequency, and monetary	K-means clustering is applied to the RFM model in various time intervals. Multiple Correspondence Analysis (MCA) is used to regularize the three dimensions of the RFM model into unified clustering centers for better business implications.	The dataset is related to telecom service and includes approximately 481,905,749 rows of data.	The study proposed a methodology utilizing the RFM model and MCA to analyze large-scale data and clustering experiments on both dimensions of RFM and time intervals. It concluded that fitting the RFM model separately in time intervals guarantees precise performance.

Table 1. Cont.

Study	Features	Methodology	Dataset	Conclusion
Abbasimehr and Shabani [15] 2019	Recency, frequency, and monetary	The Ward clustering method was used with four similarity measures including Euclidean distance, COR, CORT, and DTW. The validity of the obtained clusters was tested using the silhouette index.	Data are derived from transactions made by B2B POS customers of a bank over seven months. There are 259,000 transactions recorded, involving 2531 customers.	The study identified clusters of CPB through time series clustering and subsequently analyzed these clusters to reveal distinctive customer characteristics that enable tailored marketing recommendations for each group.
Guney et al. [16] 2020	Length, recency, frequency, monetary, and product category	This study employs the LRFMP model to categorize customers, standardizes the data using min-max normalization, and identifies customer segments using K-means clustering. Then, it employs the Apriori algorithm for association rule mining to analyze content preferences and determine customer preferences.	The dataset is sourced from an Internet protocol television operator and comprises STB-based data from 195,493 subscribers over two years.	Customers were categorized into four groups and their rental preferences were determined. The study suggests that the proposed approach effectively identifies customer groups, aiding in the development of suitable marketing strategies for implementation.
Abbasimehr and Shabani [17] 2021	Recency, frequency, and monetary	Agglomerative hierarchical clustering with Ward's method is utilized, employing various distance measures such as Euclidean distance, CORT, DTW, and CID. Additionally, traditional time series forecasting methods like ARIMA, SMA, and KNN are employed to predict customers' future clusters.	The dataset includes eleven months of customers' POS transactions from a bank.	The proposed method for segment-level customer behavior forecasting surpasses all other individual forecasters in symmetric mean absolute percentage error (SMAPE). This approach is designed to enhance customer relationship management.
Gala, et al. [18] 2022	Behavioral and purchase features (not explicitly stated)	Hierarchical clustering is utilized to segment customers according to their purchasing behaviors, with parameter linkage and the number of clusters adjusted to optimize performance. The silhouette score serves as a metric for assessing the quality of the clusters.	It comprises data from 1659 customers in the food sector, encompassing details on 5685 orders placed by these customers.	They proposed a system to automate the clustering process. It was tested on two different datasets from the supermarket and restaurant industries and achieved a silhouette score of 0.69, which indicates high accuracy in segmenting customers.

Table 1. Cont.

Study	Features	Methodology	Dataset	Conclusion
Abbasimehr and Sheikh Bagheri [19] 2022	Time series features, including measures of central tendency, variability, etc.	It proposes a new method combining Laplacian feature ranking and hybrid Support Vector Regression for customer behavior analysis. The approach is evaluated across four clustering algorithms including k-medoids, K-means, Fuzzy c-Means, and Self-Organizing Maps.	The data comprises eleven months of POS transactions in various guilds, such as home appliance stores, grocery shops, and supermarkets.	The optimal clustering algorithm varied across datasets, with K-means being most effective for grocery and k-medoids performing best for home appliance and supermarket data. In forecasting accuracy and mean SMAPE, it outperformed other methods in certain clusters and showed superior forecasting accuracy in grocery and supermarket datasets.
Abbasimehr and Bahrini [20] 2022	Recency, frequency, and monetary	Four distance measures of DTW, CID, CORT, and SBD and three clustering algorithms of hierarchical, spectral, and k-shape are utilized to identify customer groups with similar behavior. The silhouette and CID indexes are employed to evaluate and select the best clustering results. Subsequently, customer segments are labeled and analyzed based on their behavior to inform marketing strategies.	It was derived from a bank over eleven months and consists of 2,156,394 transactions conducted by consumers across grocery and appliance retailers' domains.	Customers were categorized into four groups including high-value, middle-value, middle-to-low-value, and low-value based on their behavior. Analyzing these segments revealed distinct behavioral patterns, providing valuable insights for designing effective marketing strategies and enhancing customer lifetime value.
Sun et al. [21] 2023	Recency, frequency, monetary	The methodology in this study utilizes the RFM model for customer classification, followed by the application of the BG/NBD model to predict purchase expectations and the gamma-gamma model to forecast consumption amounts. Customer lifetime value (CLV) is then calculated based on these predictions.	The dataset used in this study is the online retail dataset, which is generated from non-store online retail transactions registered in the UK.	The research concluded that the proposed method improves the accuracy of customer value, user-level correlation analysis, and explanation of intermediary effects. It can provide marketing strategies for diverse customer segments while ensuring the quantity and quality of these groups.

As is evident from Table 1, the existing literature on dynamic customer segmentation using time series features predominantly concentrates on predicting customer behavior and analyzing purchasing patterns within distinct segments, where marketing strategies are designed for each group of customers based on shared characteristics in each segment. Our study shifts the focus to analyzing each customer separately, allowing businesses

to assess and influence the purchasing behavior of their customers individually. This approach enables the design of personalized marketing strategies at the individual level, rather than the segment level, which can more effectively foster desired customer behaviors. Moreover, existing studies often overlook the transitions of customers between segments over time, while our study addresses this gap by analyzing how customers transition between segments and how CPB features should be affected by strategies for facilitating these transitions or maintaining a customer within the desired segment. In this regard, we introduce counterfactual analysis as a novel concept to the existing literature. While many studies contribute theoretical insights, fewer offer practical tools that businesses can readily implement. Our research aims to develop an efficient model that businesses can use to predict their customer behavior and assess the impact of various marketing strategies on each customer over time. This practical application should be computationally efficient to handle large datasets and is designed to enhance marketing precision and foster customer engagement and retention.

To achieve the above-mentioned objectives, our research approach begins with conducting time series clustering of customer behaviors, employing an extended version of the RFM model, and selecting an appropriate clustering algorithm based on the evaluation results. Subsequently, we employ classification algorithms to train a predictive model and forecast customer behavior status. Finally, we utilize counterfactual analysis to provide a tool that assists decision-makers in evaluating the effect of potential targeted strategies on each customer individually, aiming to retain existing customers and guide them into desired segments.

The remainder of the paper is organized as follows: Section 2 begins with an examination of algorithms suitable for time series customer segmentation. In Section 3, the proposed methodology and framework are presented. Section 4 is dedicated to providing the experimental results and Section 5 presents discussion and managerial implications. In Section 6, we draw our study to a conclusion, and finally, in Section 7, the limitations of this study and future research directions are elaborated.

2. Time Series Segmentation Algorithms

In this section, a variety of clustering algorithms well suited for time series segmentation are explored. Clustering algorithms can be broadly categorized into two groups: hierarchical and partitioning algorithms [33]. Other studies also identified grid-based, model-based, density-based, and multi-step clustering algorithms as the primary categories of clustering methods [34]. In the following, we delve into the key aspects of applying each clustering group to time series data.

Hierarchical clustering is a versatile approach in cluster analysis used to create a hierarchy of clusters using agglomerative (bottom-up) or divisive (top-down) algorithms [35]. It generates nested clusters by considering pairwise distances between data points. This is accomplished by utilizing a proximity measure referred to as the linkage metric. Hierarchical clustering can be well suited for applications where the number of clusters is challenging to define, and it can handle time series of unequal lengths when equipped with elastic distance measures like Dynamic Time Warping (DTW) [36,37]. Bekhin (2006) noted that hierarchical clustering algorithms that rely on linkage metrics face challenges related to their time complexity. Therefore, it is typically not well suited for efficiently handling large time series datasets [38]. This limitation stems from its quadratic computational complexity, meaning that the algorithm's processing time increases quadratically with the number of data points. Ward's clustering method is an agglomerative clustering algorithm, distinct from the linkage metric approach introduced by Ward in 1963. Instead, it is founded on the objective function of K-means, with the merging decision contingent on its impact on this function. It is worth noting that this clustering method is best suited for quantitative variables, as it may not be as well suited for binary variables [39]. Taking computational complexity into consideration within the broader context of clustering categorizations, it is

well established that the hierarchical clustering method is characterized by a complexity of $O(n^2)$ [39], which makes it less suitable for time series clustering applications.

Partitioning algorithms are used for grouping similar data points into distinct clusters or partitions [34]. The dataset containing n objects is progressively divided into a predefined number (k) of separate subsets through an iterative process aimed at optimizing a specific criterion function [40]. One of the widely employed algorithms in partitioning clustering is K-means [41]. It aims to create clusters by minimizing the total distance between objects within each cluster and their respective prototypes, which are typically the mean values of cluster objects [41]. In contrast, k-medoids (Partitioning Around Medoids—PAMs) assign the center of each cluster as one of the data points within the cluster, specifically the one that is closest to the other points in that cluster [35]. However, one significant challenge in both K-means and k-medoids is the need to pre-assign the number of clusters (k), which makes it challenging for the clustering of time series data [42,43]. Fuzzy clustering methods, such as Fuzzy c-Means (FCMs) and Fuzzy c-Medoids, offer a “soft” approach to clustering, where each object holds a degree of membership within each cluster [44–46]. The computational complexity of partitioning algorithms is $O(n)$ [39], which is significantly more efficient than that of hierarchical algorithms, where the complexity is $O(n^2)$, making them a better option for time series data.

The other clustering models discussed in previous studies include model-based clustering, which seeks to recover the original model from a given dataset by assuming a model for each cluster and finding the best fit of data to that model [34]. This approach involves the selection of centroids and the addition of noise with a normal distribution, which results in the recovery of a model that defines the clusters [47]. Typically, model-based methods use statistical or neural network approaches. For instance, a Self-Organizing Map (SOM), a model-based clustering method based on neural networks, has been applied in time-series clustering [48]. However, the SOM method struggles with time series of unequal lengths due to the need to define the dimension of weight vectors [35]. Model-based clustering has two main drawbacks: the need to set parameters and the reliance on user assumptions, which can lead to inaccurate cluster results and slow processing times, especially when using neural networks on large datasets [49]. Second, density-based clustering identifies clusters as subspaces of dense objects separated by subspaces of low-density objects [34]. One well-known algorithm based on the density concept is DBSCAN, which expands clusters if their neighbors are dense [50]. Chandrakala and Chandra propose a density-based clustering method in kernel feature space for clustering multi-variate time series data with varying lengths [51]. They also introduce a heuristic method for finding the initial values of the algorithm’s parameters. However, density-based clustering is not widely employed for time series data clustering due to its relatively high complexity [34]. Grid-based clustering methods discretize the data space into a finite number of cells organized in a grid-like structure. Subsequently, clustering operations are carried out on these grid cells. Two prominent examples of clustering algorithms that adopt this grid-based approach are STING [52] and Wave Cluster [53]. Finally, multi-step clustering involves the fusion of various techniques, often referred to as a hybrid method that is applied to enhance the overall quality of cluster representation [54,55].

3. Materials and Methods

A detailed, step-by-step methodology for segmenting and analyzing customer purchasing behavior utilizing the concept of counterfactual analysis is provided. Subsequently, the proposed methodology will be implemented using real-world data in the following section.

3.1. Data Understanding and Preprocessing

The dataset employed in this study consists of transactions conducted by customers of an information technology company. The dataset covers a substantial five-year period, ranging from May 2017 to March 2023, incorporating a total of 271,152 transactions attributed to 10,393 unique customers. The essential steps employed to transform the initial

data into a clean, structured, and analytically ready format are detailed in the following. The goal is to enhance the quality, consistency, and usability of the dataset, establishing a solid foundation for our study. The remainder of this section covers the handling of missing values and inconsistencies, outlier detection, feature engineering, and normalization. In the first step, duplicate records and missing values are removed. Then, data are verified against accurate formats and business rules and prepared for the next step.

Dealing with Outliers in Raw Data: Since our dataset comprises time-stamped customer transactions, with each entry structured as CustomerID, InvoiceNumber, TransactionDate, and Amount, in the first step, we focused on detecting outliers within the “Amount” field. Based on the numeric nature of the “Amount” field, we took advantage of a combined approach using the K-means algorithm as a distance-based algorithm and Z-scores to detect the outliers. To compute the Z-score for an individual transaction’s amount (let us call it x_i), Equation (1) is applied.

$$Z_i = (x_i - \mu) / \sigma, \quad (1)$$

where μ represents the mean (average) of all transaction amounts and σ is the standard deviation.

The Z-score, as depicted in Equation (1), represents a statistical measure of a score’s correlation with the mean within a set of scores which is used for outlier detection [56]. The primary assumption underlying this rule is that the variable X follows a normal distribution, leading to the Z-score having a standard normal distribution. Several studies demonstrate that the Z-score, in combination with a distance metric, can effectively detect outliers [57]. To identify outliers using the Z-score method, a predetermined threshold is typically set. It is a common practice to set the threshold to 3, but it is important to note it depends on the normal distribution of the data and also may vary depending on the context and the dataset. To suppress the Z-score assumption and since every customer’s data are valuable from an analytics perspective, a combination of the K-means algorithm with the Z-score method is employed to achieve precise and tailored outlier detection. We employed the Elbow or Within-cluster Sum of Square (WSS) method to determine the ideal number of clusters (k) for K-means clustering [58]. The result indicated that the optimal value for K lies beyond or is equal to 3. Consequently, we further explored clusters with k values ranging from 3 to 9. For a detailed breakdown of the number of samples in each cluster, please refer to Table 2.

Table 2. K-means clustering results for outlier detection: distribution of samples across clusters.

Number of Clusters	Number of Samples in Each Cluster
3	[[271,125], [10], [17]]
4	[[271,122], [7], [16], [7]]
5	[[271,121], [3], [16], [7], [5]]
6	[[271,113], [3], [10], [7], [5], [14]]
7	[[271,105], [3], [10], [7], [5], [14], [8]]
8	[[271,109], [3], [10], [7], [11], [2], [7], [3]] ¹
9	[[270,529], [3], [10], [7], [11], [2], [7], [3], [580]]

¹ The identified outliers are presented in bold font and underscored.

Alternatively, an assessment of the outliers while varying the threshold values in the Z-score method was conducted. After analyzing the results, both the K-means algorithm and the Z-score method revealed a consensus on identifying outliers when K was set to 8 and the Z-score threshold was set to 6. This agreement led to the removal of 43 outliers from the dataset. It is worth noting that when the threshold was set within the range of 3 to 5, a total of 47 outliers were detected at most. This approach can serve as a practical solution for determining the threshold of the Z-score method based on the unique characteristics of the dataset.

Feature Extraction: The popular RFM model was employed to represent customer behavior [5,7] in time intervals. The RFM model was introduced by Arthur Hughes in the 1990s to calculate the value of customer behavior. RFM encompasses recency, which

offers insights into the time that has passed since the customer's last purchase, frequency, which quantifies the total number of purchases made and reflects customer loyalty within a specified timeframe, and monetary, which pertains to the average amount expended by the customer within a time interval [59]. Subsequent studies extended the understanding of customer characteristics by introducing the length of customer engagement (i.e., [16,60–64]) as an important feature in capturing customer behavior. In this study, an additional variable of the RFM model labeled L is utilized, which signifies the lifespan of each customer in the business. Therefore, the RFML dynamic features are defined for each customer as follows (refer to Appendix A for the details of the algorithm):

- Recency (R) represents the number of days which have passed since the last purchase prior to the current time interval;
- Frequency (F) pertains to the total number of purchases within the current time interval;
- Monetary (M) signifies the total purchasing amount during the current time interval;
- Lifespan (L) reflects the number of days which pass between the initial purchase and the last one in the current time interval.

This transformation results in the conversion of raw customer transaction data into RFML feature vectors that are calculated on time intervals. In determining the length of the time interval, we consider various factors, including the nature of the data, specific domain requirements, and the objectives of our analysis. Our goal is to strike a balance between capturing significant patterns within each segment without overly fragmenting the data. After careful consideration of the specified criteria, insights from the existing literature, and consultations with domain experts, a one-month time interval was selected for this study. Consequently, we extracted 37,442 samples that encapsulate the RFML dynamic features of customer behavior from the raw transactional data. In our subsequent data refinement process, we subjected each RFML feature to the outlier detection method which was used on raw data as well. Consequently, we identified and removed a total of 764 customer behavior samples that exhibited outlier characteristics. The min-max normalization technique to normalize the features, scaling them into the [0,1] range, was utilized in this part [63].

3.2. Customer Behavior Segmentation

Clustering is a complex task, where the quality of outcomes relies significantly on two critical decisions: the selection of an appropriate clustering algorithm and the choice of a suitable distance measure [64]. Considering the previously discussed benefits and drawbacks of various clustering methods within the context of time series data and the time complexity associated with each method, this study aims to assess the practical utility of the commonly employed partitioning clustering technique, K-means, in the analysis of customer behavior dynamics. Given that the K-means algorithm relies on the optimization criterion involving the distances between data points, our focus is on the examination of various distance metrics to determine the most effective one in this context. In particular, our goal is to employ a variety of distance metrics, encompassing both commonly used metrics in clustering problems and those specifically designed for time series data. Furthermore, the research seeks to incorporate a multi-step clustering approach, guided by the dual objectives outlined in the following section, to gain a deeper understanding of the dynamics of customer behavior.

3.3. Customer Behavior Status Prediction

In this study, a classification algorithm is employed to predict the status of new customer behaviors as represented by the RFML feature vectors. This approach encompasses several distinct objectives. Firstly, the segment or status membership of new customer behaviors is determined using classification algorithms. A multi-step clustering methodology, encompassing both clustering and classification algorithms, is employed to enhance the overall performance of cluster prediction. Secondly, a feature importance analysis is conducted to identify the key features that make significant contributions to each segment,

thereby providing valuable insights to decision-makers. Lastly, a tool inspired by counterfactual analysis is devised, empowering decision-makers to evaluate the consequences of altering individual features on the status of individual customers within a segment, as opposed to considering the entire cluster as a whole. To achieve this, a comprehensive evaluation of three classification algorithms—Random Forest, logistic regression, and Decision Tree algorithms—is undertaken. The purpose of this examination is to identify the most appropriate algorithm aligned with our research objectives. In our selection process, three fundamental criteria are established, guided by the aforementioned goals that the chosen algorithm must meet. Firstly, non-distance-based algorithms are employed to complement the distance-based clustering approach applied in the preceding step. Secondly, the algorithm is expected to yield comprehensive insights into feature importance. Lastly, the ability to provide probabilities associated with each class membership is another crucial criterion under consideration.

3.4. Counterfactual Analysis and Personalized Strategies

In this section, we aim to take advantage of the concept of counterfactuals. This term, with its roots in the works of philosophers David Hume and John Stewart Mill, has acquired computer-friendly semantics in recent decades. A common query within the counterfactual realm necessitates retrospective reasoning, often posed as, “What if I had acted differently?” In fact, counterfactuals are the building blocks of scientific thinking as well as legal and moral reasoning [64]. This section employs the exploration of counterfactuals in our study.

In the domain of counterfactual analysis, we come across expressions represented as $P(y|x, x', y')$. These expressions symbolize the probability of event Y taking a specific value y under the condition that X was x , assuming we have observed X as x' and Y as y' . To illustrate this with an example, consider the probability that a customer's segment would be y “=Loyal Patron” if the number of his monthly purchases (x') grew 10 percent for 3 consecutive months, given that his actual segment is y' “=at Risk of Losing” and the number of his monthly purchases is x' . This framework allows us to explore hypothetical scenarios and evaluate the likelihood of outcomes when certain variables or conditions are altered, based on the observed data and relationships between variables.

These statements, capturing counterfactual probabilities, are computable when we have access to functional or structural equation models or related properties of such models. In other words, as expounded by Pearl in his work [65], these models provide the necessary framework for quantifying and reasoning about such counterfactual scenarios. They enable us to explore what might have occurred under different circumstances, given the observed outcomes and the underlying structural relationships between variables. In this study, the aim is to harness the potential of counterfactual analysis and provide the basis of a practical tool that empowers businesses to assess a multitude of scenarios. This approach enables the identification of highly personalized marketing proposals for individual customers, unlike traditional generalized strategies employed for each customer group. To do so, we will use the possibility provided by the algorithms chosen in the prior phase and extract the influence of each feature, thereby contributing to the comprehensive understanding of customer behavior determinants. The objective of this approach is to enhance decision making processes within the domain of customer-centric marketing strategies. Figure 1 provides an overview of the proposed methodology designed to address the objectives of this study and will be applied to real-world data in the subsequent section.

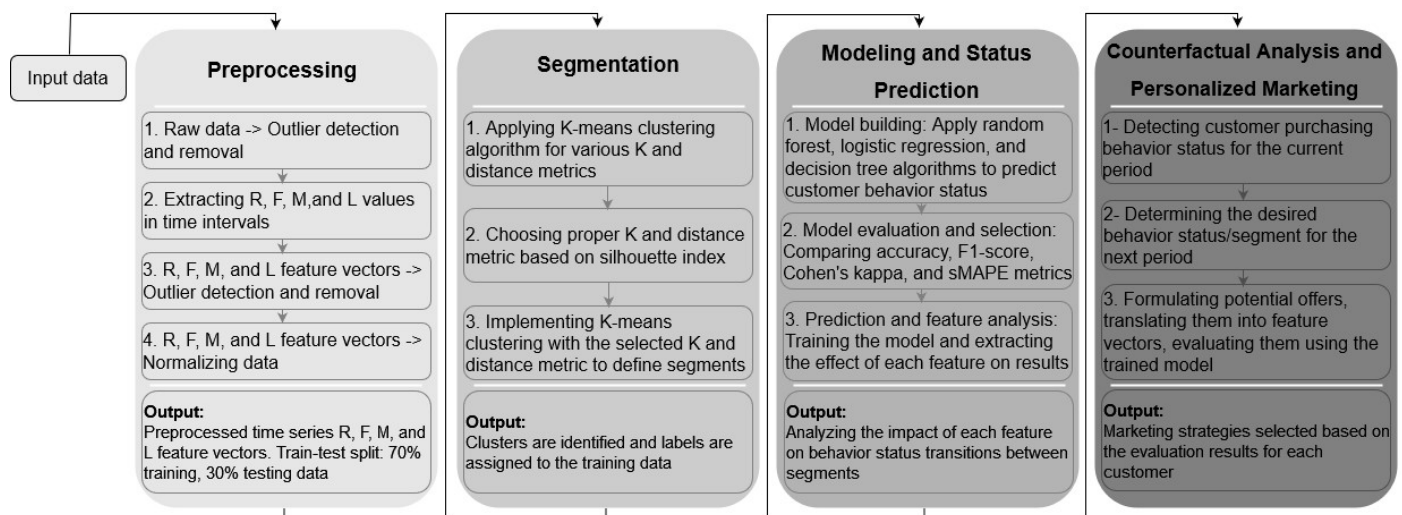


Figure 1. Proposed methodology for dynamic customer behavior analysis and evaluating personalized marketing strategies.

4. Results

This section presents the outcomes of applying the methodology introduced in the previous section to real customer transaction data, illustrating the practical implementation of our approach. Following the preprocessing phase as detailed in the preceding section, we proceed with a step-by-step process involving customer behavior segmentation, customer behavior status prediction, and the application of counterfactual analysis. This analysis evaluates the impact of potential marketing strategies on customer status, helping to determine tailored strategies for each customer.

4.1. Customer Purchasing Behavior Segmentation

In this step, the K-means algorithm is applied to the time series RFML feature vectors using various distance metrics. Since the K-means algorithm of the Scikit-learn library primarily relies on the Euclidean distance metric for clustering, to assess the performance of the K-means algorithm with various distance metrics, we implemented the K-means algorithm using Python version 3.11.5 within the Anaconda software version 23.7.4. Subsequently, the results were analyzed using the silhouette index [66–68] to ascertain which distance metric yields denser and distinctly separated clusters of customer behaviors, represented by the time series RFML feature vectors. Since the distance measure has a direct impact on the clustering quality of time series data [69], we applied Euclidean, Manhattan, and Chebyshev distance metrics, which are commonly used in clustering [39], and Dynamic Time Warping (DTW) [70], temporal correlation coefficient (CORT) [71,72], complexity-invariant distance (CID) [73], designed for time series data. The results are shown in Table 3.

Table 3. K-means clustering with a varied number of clusters and distance measures: silhouette results.

K	Euclidean	Manhattan	Chebyshev	CORT	DTW	CID
4	0.461	0.457	0.426	0.114	0.450	0.283
5	<u>0.486</u> ¹	0.485	0.382	0.206	0.392	0.200
6	0.427	0.458	0.425	0.205	0.373	0.227
7	0.469	0.461	0.446	0.151	0.385	0.227
8	0.438	0.470	0.404	0.122	0.417	0.221
9	0.447	0.427	0.416	0.128	0.428	0.215

¹ The optimal result for each metric is presented in bold font, while the best result across all metrics is additionally underscored for emphasis.

The results demonstrate that both Euclidean and Manhattan distances exhibit superior performance in partitioning customers' behavior data into five distinct behavioral groups. This suggests that these distance metrics are particularly effective in distinguishing the dynamics of customers' behaviors represented by RFML feature vectors, with the Euclidean distance showing slightly better performance. The result of the optimal number of clusters was entirely consistent with using the WSS method to determine the optimal k value for K-means. Table 4 presents the cluster statistics, encompassing cluster size and cluster compactness, calculated as the Within-cluster Sum of Squares.

Table 4. Analyzing cluster size and compactness in K-means using Euclidean distance.

Cluster	Size	Cluster Compactness
0	1264	63.98
1	13,552	116.02
2	2731	135.06
3	1443	210.41
4	6684	108.67

Figure 2 presents both pairwise feature analysis and a 3D visualization of the clusters. The pairwise analysis demonstrates how the segments manifest from the viewpoint of each feature pair. Notably, better segment separation is observed when examining data through the lens of lifespan, which can show the importance of this feature in CPB clustering. By integrating all plots from the pairwise analysis, we gain insights into the overall data and cluster structure, aligning with the presentation in the 3D plot in Figure 2. Following this phase, it is crucial to meticulously analyze the outcomes of each cluster based on business rules and objectives. Figure 3 illustrates the distribution of each feature within the clusters, allowing for a comparison with the overall mean across all clusters. This visualization provides valuable insights into the positioning of features within clusters and aids in extracting shared customer behaviors within each group. Cluster 0 belongs to dormant customers. Customers in this segment exhibit very high recency and very low frequency and monetary value, with a slightly above-average lifespan. This indicates that these customers have not made recent purchases, purchase very infrequently, and spend very little when they do. They may have been active in the past but churned for a while and are disengaged. Infrequent low-spenders formed cluster 1. This cluster is characterized by low recency, frequency, and monetary value and a short lifespan. These are relatively new customers who have made recent purchases but buy infrequently and spend little. Their engagement is minimal and spread over a short period. Long-term loyal customers are in cluster 2. They show moderate recency and relatively high frequency and monetary value, with a very long lifespan. They are consistent purchasers who engage regularly and spend a good amount over an extended period. Their moderate recency suggests they continue to engage with the business, making them valuable in the long term.

Cluster 3 includes high-value shoppers. This segment stands out with very low recency, very high frequency, and the highest monetary value, coupled with a long lifespan. These customers make recent, frequent purchases and spend significantly, representing the most valuable customers for the business. The last cluster, cluster 4, belongs to moderate-value shoppers. Customers in this cluster have moderate recency and relatively moderate frequency and monetary value compared to other clusters, with an above-average lifespan. These customers have been with the business for a considerable time but are not the most frequent or high-spending buyers. They represent a segment of moderately valuable customers who could be encouraged to increase their engagement and spending through targeted promotions and personalized marketing strategies. These results enhance our understanding of common purchasing behaviors within each group and provide valuable insights for studying and evaluating customer transitions between segments.

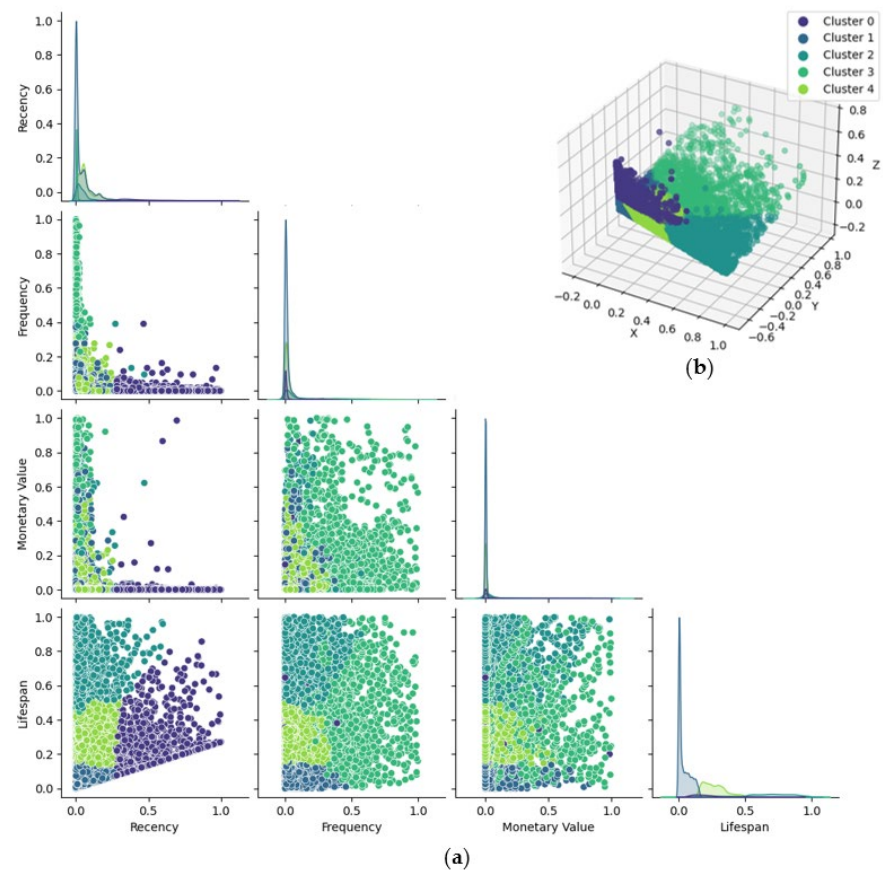


Figure 2. Visual exploration of customer behavior clusters: (a) pairwise feature analysis; (b) 3D visualization.

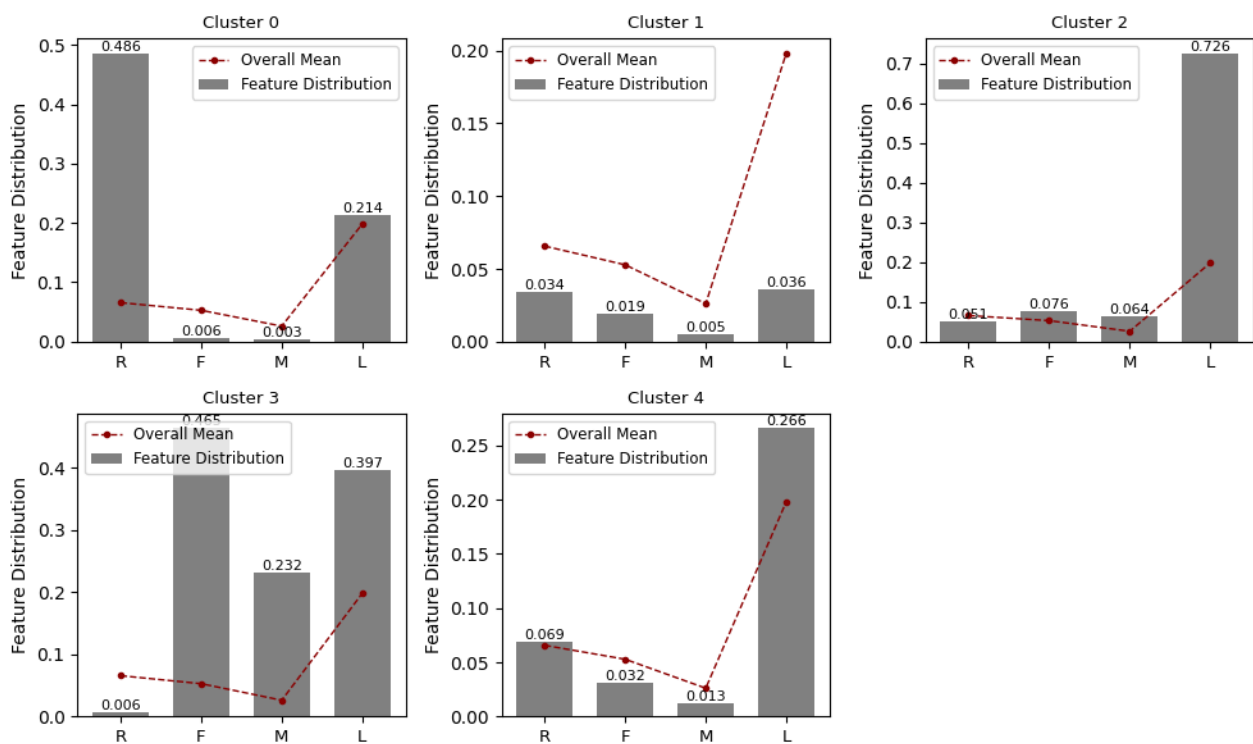


Figure 3. Analyzing feature distributions across clusters compared to overall mean values: a visual examination in our case study.

4.2. Customer Purchasing Status Prediction

The Random Forest, logistic regression, and Decision Tree algorithms are utilized to predict customer behavior status. The goal is to train these algorithms using cluster labels extracted from the K-means algorithm. To identify optimal parameters for each algorithm, 5-fold cross-validation is employed. The performance of each algorithm in assigning status to each customer behavior is then assessed and compared using standard metrics including accuracy, F1 score, and Cohen's kappa as commonly used validation metrics to evaluate the performance of multi-class classification models [74]. Additionally, the symmetric mean absolute percentage error (sMAPE) metric is utilized, calculated using Equation (2), as the main metric used to assess forecasting accuracy in time series competitions [75]. The results are presented in Table 5, where y_i and \hat{y}_i are the actual and predicted i th values, respectively.

$$\text{sMAPE} = \frac{100}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{\frac{|y_i| + |\hat{y}_i|}{2}}, \quad (2)$$

Table 5. Performance in predicting customer behavior status: evaluation metrics comparison.

Classifier	Accuracy	F1-Score	Cohen's Kappa	sMAPE
Random Forest	0.969	0.953	0.982	2.821
Logistic Regression	0.998 ¹	0.999	0.999	0.345
Decision Tree	0.986	0.959	0.981	0.361

¹ The optimal result for each metric is presented in bold font, while the best result across all metrics is additionally underscored for emphasis.

The results indicate that both logistic regression and Decision Tree algorithms exhibit strong performance in predicting customer behavior. However, logistic regression consistently outperforms the alternative algorithm across all metrics, demonstrating superior predictive capabilities. Following that, the impacts of each feature on the prediction are derived from the trained logistic regression model, providing crucial inputs for the subsequent counterfactual analysis.

4.3. Counterfactual Analysis and Personalized Strategies

In this phase, the primary objective is to assess the influence of each feature on customers transitioning between different segments or statuses. By identifying key features driving CPB, businesses can tailor marketing strategies to individual customers and potentially guide them towards desired outcomes if necessary. This approach not only has the potential to increase the likelihood of success in marketing endeavors but also holds the promise of increasing revenue for companies. Therefore, our ultimate goal is to provide companies with a reliable tool to evaluate the impact of each offer on the customer's purchasing behavior before its implementation. The information derived from predicting the customer's purchase status, which shows the importance of each feature in the forecasted scenario, lays the essential groundwork for this part. The equation derived from the logistic regression model, as depicted in Equation (3), enables the assessment and quantification of the probability of a future behavior aligning with each segment.

$$P(Y = c | X = x; \omega) = \frac{\exp(\beta_c^T x)}{\sum_{n=1}^C \exp(\beta_n^T x)}, \quad (3)$$

Suppose we have a multi-class problem with C classes ($C \geq 2$). Equation (3) provides the probability that a new observation, denoted as x , belongs to class c , with $\omega = [\beta_1^T; \beta_2^T; \dots; \beta_C^T]$, $\omega \in \mathbb{R}^{(C)d}$ being a collection of the different parameter vectors of C linear models. The parameter vectors of the models are employed to assess the influence of each feature on the customer status and CPB transitions, as illustrated in Figure 4. This

equation serves as a basis for developing a framework to forecast the effects of different potential strategies on customer purchasing behavior. It provides valuable insights into guiding customers to transition from a suboptimal cluster, which may not align with the business's objectives. Adjusting behavior features through marketing offers can facilitate this transition toward a desired cluster. By formulating the effect of potential marketing scenarios on R, F, M, and L parameters for each customer, businesses can evaluate the impact of these scenarios on customer purchasing behavior before their implementation and predict customer status. To demonstrate the applicability of our framework, let us delve into the transition from cluster 0, representing dormant customers, to cluster 3, comprising high-value shoppers. As depicted in Figure 4, the frequency of purchases exerts the most significant influence on this transition. Following closely is the reduction in the time intervals between purchases. To facilitate such transitions effectively, the company should implement strategies that encourage customers to make more frequent purchases distributed in shorter time intervals. In this scenario, emphasizing a higher transactional monetary value has a comparatively less potent impact on this transition.

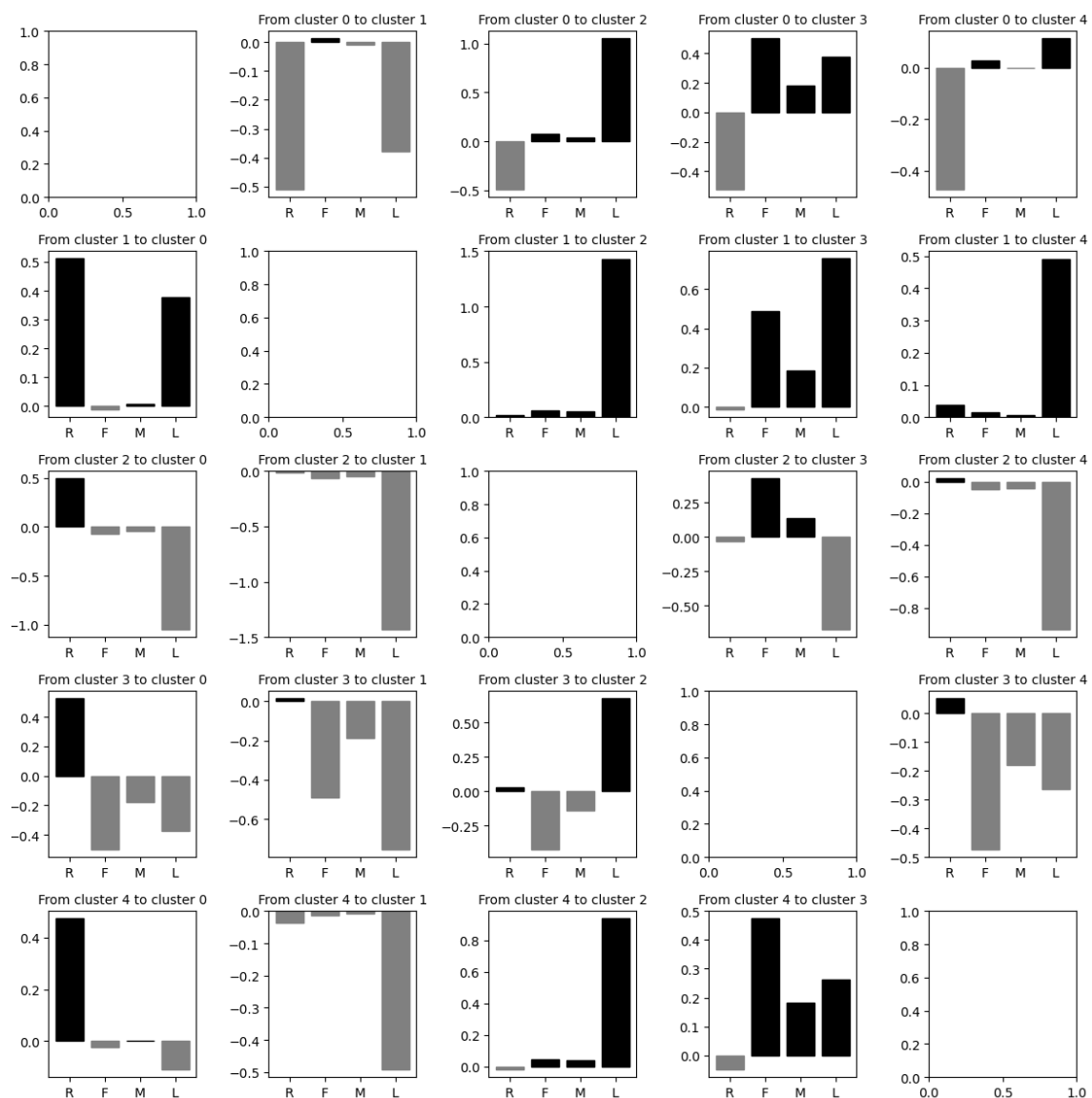


Figure 4. Analyzing the impact of each feature on customer transition: insights from the logistic regression predictive model.

Table 6 demonstrates the results of implementing some potential scenarios on customer behavior #1944 in the dataset under study. The current behavior of this customer falls under cluster 0. Analyzing the feature distribution of this cluster, as depicted in Figure 3, revealed this cluster comprises customers who had churned and did not make a purchase in a while (approximately a year) but have recently re-engaged with a new purchase. This customer behavior is analyzed to determine the potential scenarios that cause the customer to transition to a more engaged cluster such as clusters 2, 3, and 4.

Table 6. Potential scenarios for transitioning customer behavior #1944 to a desired cluster.

Potential Scenarios	Current Status	Desired Status		
	Cluster 0	Cluster 2	Cluster 3	Cluster 4
Current Status	X			
1. Making a purchase within the next 3 months with the current M value ($R\downarrow$)	X	—————→	X	
2. Increasing F by a factor of 4 and making purchases within the next two months ($R\downarrow F\uparrow$)	X	—————→	X	
3. Make purchases with current values of F and M every two months for two years ($R\downarrow L\uparrow$)	X	————→	X	

↓ (Down arrow) indicates a decrease or reduction in the value of the corresponding feature. ↑ (Up arrow) indicates an increase or augmentation in the value of the corresponding feature.

To initiate the transition process, employing insights from Figure 4, hypothetical scenarios for this customer should be devised which aim at guiding them into the desired clusters. Subsequently, for each scenario, the changes induced by the scenario are translated into their R, F, M, and L features. In the following step, the question “What if the customer behaves like this?” is addressed to assess the effect of the scenario on the subsequent state of the customer. These hypothetical scenarios can be evaluated using the trained logistic regression model to forecast the next status of the customer if the strategy is implemented. Table 6 presents three scenarios with different behaviors and their impact on the customer’s transition. In the first scenario, the customer makes a purchase within the next three months, reducing their recency while keeping their current monetary value unchanged. Figure 4 indicates that a reduction in recency has the most significant effect on the customer’s transition from cluster 0 to cluster 4. This decrease in the recency of CPB #1944 would transition it from a dormant customer to a moderately engaged one. In the second scenario, by providing proper marketing offers, we need to increase the purchase frequency by a factor of four, and the customer should make purchases within the next two months. This needs a significant change in customer behavior, and marketing experts can devise a step-by-step transition through this segment and turn the customer into a high-value shopper belonging to cluster 3.

In the third scenario, the customer maintains the current frequency and monetary value but makes purchases every two months over the next two years. This consistent engagement reduces recency, and after two years, increasing the lifespan, CPB #1944 transitions to cluster 2 to become a long-term loyal customer. The result facilitates the decision making process for determining effective marketing propositions to offer to the customer. The aim is to expedite the transition of customers to a desired cluster or state, ensuring optimal outcomes in terms of speed and efficacy.

To further explore the application of this methodology, Table 7 examines the impact of different scenarios on CPB #3309, currently categorized under cluster 3 as a high-value shopper. Despite them being a valuable customer, as revealed in the results presented in Figure 3, we aim to investigate scenarios that could keep the customer in the current status or potentially lead to a transition into an undesired status of clusters 0, 2, and 4 for this

customer. This information can assist decision-makers in developing proactive strategies to prevent the customer from moving out of a profitable cluster.

Table 7. Potential scenarios for maintaining customer behavior #3309 in the desired cluster and their outcomes on customer status.

Potential Scenarios	Current Status	Undesired Status		
	Cluster 3	Cluster 0	Cluster 2	Cluster 4
Current Status	X			
1. Keep the same metrics for the next two years	X			
2. Inactive for three months	X			
3. Inactive for seven months (R↑)	X →	X		
4. Decreasing F by a factor of two (F↓)	X →		X	
5. Reduce F by a factor of two after two years (F↓L↑)	X →	X		

↓ (Down arrow) indicates a decrease or reduction in the value of the corresponding feature. ↑ (Up arrow) indicates an increase or augmentation in the value of the corresponding feature.

The first scenario indicates that maintaining the current metrics of recency, frequency, and monetary value for CPB #3309 over the next two years does not change the customer's current status. The second and third scenarios examine the impact of purchase recency alterations on customer status. The results show that if this customer becomes inactive for three months, the status remains unchanged. However, seven months of inactivity causes the customer to shift to cluster 0, which belongs to dormant customers. The last two scenarios indicate that if the purchase frequency decreases by half in the short term, the customer will move to cluster 4, which includes moderate-value shoppers. Conversely, if this reduction happens gradually over two years, the customer shifts to cluster 2 and remains loyal, maintaining good frequency and monetary value. Nevertheless, they will no longer be categorized as a high-value customer in cluster 3. These assessments enable businesses to identify which changes in customer purchasing behavior will affect a customer's current status within a desired cluster. For instance, while the period of inactivity may be manageable for the business if it spans three months for this customer, extending it to more than three months signifies a potential churn risk and leads the customer to cluster 0. In such cases where transitions are suboptimal for a business, strategic marketing initiatives can be devised to steer the customer back toward favorable engagement. Figure 5 provides further insights into the effort required for transitioning between different segments. It is calculated based on coefficients assigned to each cluster obtained from the prediction model.

The analysis of Figure 5 reveals that transitioning from cluster 1 (infrequent low spenders) to cluster 2 (long-term loyal customers) poses the most significant challenge. Conversely, it suggests that moving between clusters 0 (dormant customers) and 4 (moderate value shoppers) appears notably more manageable.

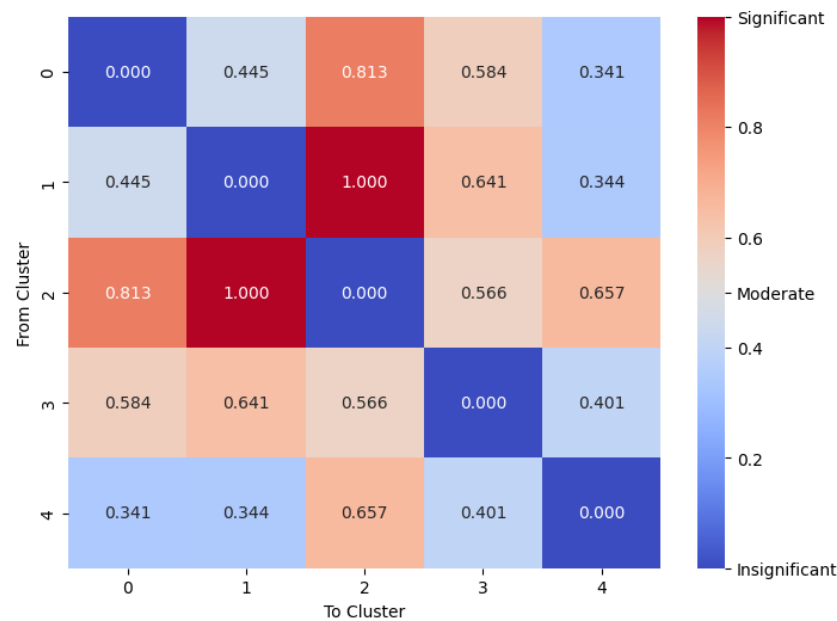


Figure 5. Effort analysis for customer purchasing behavior transitions.

5. Discussion and Managerial Implications

Traditional “static segmentation” methods capture customer behavior at a single point in time and treat extended periods as a single snapshot. This approach overlooks the dynamic changes in customer behavior over time, limiting its ability to predict future purchasing behaviors. Our dynamic segmentation approach, on the other hand, captures customer behavior in shorter, more frequent time intervals. This method allows for the tracking of evolving patterns and training predictive models to forecast customer purchasing behavior over time. Consequently, it provides a more dynamic and comprehensive analysis of customer interactions.

By shifting the focus from segment-level to individual customer behavior analysis and leveraging time series segmentation and counterfactual analysis, we proposed a framework that provides a basis for more precise and personalized marketing efforts. To capture the dynamics of customer purchasing behavior, we first represented customer behavior using time series feature vectors and then conducted dynamic segmentation. In the second part, a logistic regression model was trained to forecast future customer purchasing behavior. Utilizing the capabilities of the trained model, we extracted the impact of each behavioral feature on customer transition between segments, which provides insights for devising personalized marketing strategies based on the customer’s current behavior. In the final part, by leveraging counterfactual analysis and the trained predictive model, we could evaluate the effect of various marketing strategies on the future behavior of the customer, which empowers businesses to implement the best strategy for each customer separately. This approach enables businesses to predict how changes in marketing strategies impact individual customer behaviors before implementation. This level of personalization allows for more effective marketing campaigns, as strategies can be tailored to the specific behavioral characteristics of each customer, thereby fostering stronger customer relationships and enhancing loyalty. The implications of this approach are manifold, ranging from improved customer retention to increased revenue through tailored marketing interventions. Moreover, this approach was designed to handle large datasets efficiently, making it suitable for real-world applications.

One of the novel contributions of our research is the focus on customer transitions between segments over time. By analyzing how customers transition between segments and identifying the features that drive these transitions, businesses can develop strategies to either retain customers in desirable segments or guide them toward more valuable segments.

This dynamic approach ensures that marketing efforts are continually adapted to the evolving behaviors and preferences of customers, thus maintaining relevance and effectiveness.

The experimental results presented in our study provide valuable insights into the effectiveness of the clustering and prediction algorithms for customer purchasing behaviors. The superior performance of the Euclidean distance metric in clustering and the logistic regression algorithm in predicting customer behavior status underscores the importance of choosing appropriate methodologies for different aspects of customer behavior analysis.

This case study was conducted on data from an information technology company that provides various packages of services, software, and hardware products. The transaction data utilized in this study allow for deriving the customer purchasing behavior characteristics required for applying the extended version of the RFM model. These types of data are commonly stored by a diverse range of businesses, including retail stores, B2B wholesalers, banks, and telecom companies. These businesses regularly capture and maintain similar transaction data, making our approach broadly applicable across various e-commerce contexts. Consequently, our methodology has the potential for wide implementation and can provide valuable insights and personalized marketing strategies for companies operating in different sectors.

From a managerial perspective, the adoption of our proposed framework requires a shift in mindset from segment-based marketing to a more individualized approach. Managers should invest in advanced data analytics tools and develop the necessary skills within their teams to analyze and interpret complex customer data. Additionally, a continuous feedback loop should be established to monitor the effectiveness of implemented strategies and make necessary adjustments based on real-time data insights. This proactive and data-driven approach to customer management can significantly enhance the strategic decision making process, leading to more successful marketing outcomes.

6. Conclusions

In the context of business and industry, the ability to forecast the impact of marketing offers on customer behavior before their implementation is crucial for guiding strategic decisions, optimizing resource allocation, and ensuring the success of marketing campaigns. In this study, we conducted time series customer segmentation, developed a predictive model to forecast customer status, and analyzed the dynamics of customer behavior at an individual level, moving beyond the segment-level approach focused on in previous studies. We addressed customer transitions between segments, an aspect previously overlooked. To achieve this objective, we employed counterfactual analysis. This approach provided the analytical basis to analyze customer transitions and examine the outcomes of different strategies on customer behavior before their implementation and design marketing actions at an individual level.

Our study began by evaluating the practical applicability of the K-means clustering technique, employing various distance metrics, encompassing both commonly used metrics and those specifically designed for time series data. Customer behavior was represented by time series feature vectors including recency, frequency, monetary value, and lifespan extracted at one-month intervals. The results showed the potential of Euclidean and Manhattan distance metrics in effectively separating customer data into distinct behavior groups, with the Euclidean distance exhibiting slightly better performance. This highlights their capacity to distinguish the patterns in time series customer behavior compared to other distance metrics utilized in this study. Subsequently, we investigated the performance of the Random Forest, logistic regression, and Decision Tree algorithms in predicting customer behavior status. Our exploration into classification algorithms highlighted the proficiency of the logistic regression algorithm in predicting customer behavior status, achieving remarkable performance metrics with an accuracy of 0.9981, F1 score of 0.999, Cohen's kappa of 0.999, and sMAPE of 0.345. Consequently, our analysis of potential scenarios involving the alteration of customer behavior characteristics through marketing offers underscores the applicability and efficacy of our framework. Our findings demonstrate

how businesses can strategically guide customers from suboptimal behavior to a targeted status, while also effectively maintaining customers in their optimal behavior. This insight is invaluable for businesses seeking to optimize their marketing strategies and establish long-term customer satisfaction and loyalty.

7. Limitations and Future Research Directions

While our study contributes valuable insights to the realm of customer purchasing behavior analytics, it is essential to acknowledge some limitations inherent in our research design. Our analysis relies on data derived from a single-time experiment conducted with a specific company. Given that effective customer relation and retention management is an ongoing endeavor necessitating continuous segmentation and state predictions, our reliance on a one-time experiment raises methodological considerations.

Adapting to the evolving landscape of customer behavior segmentation and predictive models ideally requires retraining on updated data. Several strategies can be implemented to maintain accuracy and relevance in forecasting customer behavior. An automated data pipeline can be implemented to ensure models are trained on the latest data, minimizing the lag between data collection and model updates. For instance, if new transaction data are collected daily, they can be automatically integrated into the training process, ensuring the models reflect the most current customer behavior. Alternatively, if immediate updates are not required or feasible, establishing a regular retraining schedule, such as monthly or quarterly, ensures models incorporate recent trends. Additionally, creating a feedback loop to compare predicted and actual behaviors can refine models for better accuracy. Consequently, by setting predefined performance thresholds, models can be retrained automatically when their performance drops below these levels. Training the model on such data may inadvertently compromise prediction quality, as these customers might have exhibited undesirable behavior but were adjusted successfully. To address this issue, previously targeted customers can be excluded from the training set or additional features should be incorporated to capture historical information about customer behavior statuses. Further research is required to address this issue.

To enhance the development of marketing strategies, a key avenue for improvement involves incorporating additional data types, specifically demographic characteristics of customers and their interactions with the business. This descriptive information can provide more comprehensive insights for developing customized marketing and retention strategies tailored to customer preferences. However, in our current setup, relying on data from the specific company, data availability constrained our ability to integrate such features. Future endeavors are needed to study such challenges and opportunities to enhance marketing and retention capabilities. Additionally, longitudinal studies that track the effectiveness of marketing strategies over extended periods can provide valuable insights into the evolution of customer behavior.

Our next step involves addressing the limitations of this study and developing a software application grounded in the mathematical insights derived from logistic regression and counterfactual analysis. This tool will incorporate a continuous training model that periodically updates vector parameters. The aim is two-fold: to provide businesses with the required foundation for evaluating various marketing offers for each customer individually and to offer a step-by-step approach, simplifying the process of transitioning customers to a targeted status from the business perspective.

Author Contributions: Conceptualization, M.E.J.; methodology, M.E.J. and A.E.; software, M.E.J.; validation, M.E.J. and A.E.; data curation, M.E.J.; writing—original draft preparation, M.E.J.; writing—review and editing, M.E.J. and A.E.; visualization, M.E.J.; supervision, A.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are unavailable due to the company's privacy policy restrictions.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Customer Purchasing Behavior Feature Extraction Algorithm

Algorithm A1: Extracting Behavioral Feature Vectors

Data: A chronologically ordered sequence of transactions such that each row contains (*CustomerId*, *InvoiceNo.*, *TransactionDate*, and *Value*) columns

Result: Feature vectors consist of R, F, M, and L values calculated for each customer in each month

```

1      Read input data in  $df_{data}$ ;
2      Extract the year and month from the date column;
3      Group  $df_{data}$  by customer, year, and month in  $df_{rfml}$ ;
4      foreach unique customer, year, and month in  $df_{rfml}$  do
5          Find the first and last purchase dates in the current month;
6          Find the most recent purchase date before the current month;
7           $R \leftarrow$  Difference in days between the first purchase date and the previous
purchase date;
8           $Temp \leftarrow$  Count and sum purchases in  $df_{data}$  within the current month;
9           $F \leftarrow$  Tempcount;
10          $M \leftarrow$  Tempsum;
11          $L \leftarrow$  Difference in days between the last purchase date in the current month and
the first purchase date for the customer;
12     end
13     return  $df_{data}$ 
```

References

- Xiahou, X.; Harada, Y. B2C E-Commerce Customer Churn Prediction Based on K-Means and SVM. *J. Theor. Appl. Electron. Commer. Res.* **2022**, *17*, 458–475. [\[CrossRef\]](#)
- Kotler, P.; Armstrong, G. *Principles of Marketing*, 11th ed.; Prentice-Hall: Upper Saddle River, NJ, USA, 2006.
- Huang, S.-C.; Chang, E.-C.; Wu, H.-H. A Case Study of Applying Data Mining Techniques in an Outfitter's Customer Value Analysis. *Expert Syst. Appl.* **2009**, *36*, 5909–5915. [\[CrossRef\]](#)
- Chang, E.-C.; Huang, S.-C.; Wu, H.-H. Using K-means Method and Spectral Clustering Technique in an Outfitter's Value Analysis. *Qual. Quant.* **2010**, *44*, 807–815. [\[CrossRef\]](#)
- Alves Gomes, M.; Meisen, T. A Review on Customer Segmentation Methods for Personalized Customer Targeting in E-commerce Use Cases. *Inf. Syst. e-Bus. Manag.* **2023**, *21*, 527–570. [\[CrossRef\]](#)
- Aksoy, N.C.; Kabadayi, E.T.; Yilmaz, C.; Alan, A.K. A Typology of Personalisation Practices in Marketing in the Digital Age. *J. Mark. Manag.* **2021**, *37*, 1091–1122. [\[CrossRef\]](#)
- Sarkar, M.; Puja, A.R.; Chowdhury, F.R. Optimizing Marketing Strategies with RFM Method and K-Means Clustering-Based AI Customer Segmentation Analysis. *J. Bus. Manag. Stud.* **2024**, *6*, 54–60. [\[CrossRef\]](#)
- Dibb, S. Market Segmentation: Strategies for Success. *Mark. Intell. Plan.* **1998**, *16*, 394–406. [\[CrossRef\]](#)
- Miguéis, V.L.; Camanho, A.S.; Cunha, J.F.E. Customer Data Mining for Lifestyle Segmentation. *Expert Syst. Appl.* **2012**, *39*, 9359–9366. [\[CrossRef\]](#)
- Safari, F.; Safari, N.; Montazer, G.A. Customer Lifetime Value Determination Based on RFM Model. *Mark. Intell. Plan.* **2016**, *34*, 446–461. [\[CrossRef\]](#)
- Manjunath, K.; Suhas, Y.; Kashef, R. Distributed Clustering Using Multi-Tier Hierarchical Overlay Super-Peer Peer-to-Peer Network Architecture for Efficient Customer Segmentation. *Electron. Commer. Res. Appl.* **2021**, *47*, 101040. [\[CrossRef\]](#)
- Calvet, L.; Ferrer, A.; Gomes, M.I.; Juan, A.A.; Masip, D. Combining Statistical Learning with Metaheuristics for the Multi-Depot Vehicle Routing Problem with Market Segmentation. *Comput. Ind. Eng.* **2016**, *94*, 93–104. [\[CrossRef\]](#)
- Murray, P.W.; Agard, B.; Barajas, M.A. Market Segmentation Through Data Mining: A Method to Extract Behaviors from a Noisy Data Set. *Comput. Ind. Eng.* **2017**, *109*, 233–252. [\[CrossRef\]](#)
- Song, M.; Zhao, X.; E, H.; Ou, Z. Statistics-based CRM Approach via Time Series Segmenting RFM on Large-scale Data. *Knowl. Based Syst.* **2017**. [\[CrossRef\]](#)
- Abbasimehr, H.; Shabani, M. A New Methodology for Customer Behavior Analysis using Time Series Clustering: A Case Study on a Bank's Customers. *Kybernetes*, 2019; ahead-of-print. [\[CrossRef\]](#)
- Guney, S.; Peker, S.; Turhan, C. A Combined Approach for Customer Profiling in Video on Demand Services Using Clustering and Association Rule Mining. *IEEE Access* **2020**, *8*, 84326–84335. [\[CrossRef\]](#)

17. Abbasimehr, H.; Shabani, M. A New Framework for Predicting Customer Behavior in Terms of RFM by Considering the Temporal Aspect Based on Time Series Techniques. *J. Ambient. Intell. Humaniz. Comput.* **2021**, *12*, 515–531. [\[CrossRef\]](#)
18. Galal, M.; Salah, T.; Aref, M.; ElGohary, E. Smart Support System for Evaluating Clustering as a Service: Behavior Segmentation Case Study. *Int. J. Intell. Comput. Inf. Sci.* **2022**, *22*, 35–43. [\[CrossRef\]](#)
19. Abbasimehr, H.; Sheikh Bagheri, F. A Novel Time Series Clustering Method with Fine-tuned Support Vector Regression for Customer Behavior Analysis. *Expert Syst. Appl.* **2022**, *204*, 117584. [\[CrossRef\]](#)
20. Abbasimehr, H.; Bahrini, A. An Analytical Framework Based on the Recency, Frequency, and Monetary Model and Time Series Clustering Techniques for Dynamic Segmentation. *Expert Syst. Appl.* **2022**, *192*, 116373. [\[CrossRef\]](#)
21. Sun, Y.; Liu, H.; Gao, Y. Research on Customer Lifetime Value Based on Machine Learning Algorithms and Customer Relationship Management Analysis Model. *Heliyon* **2023**, *9*, e13384. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Wang, C.-H. Apply Robust Segmentation to the Service Industry Using Kernel-Induced Fuzzy Clustering Techniques. *Expert Syst. Appl.* **2010**, *37*, 8395–8400. [\[CrossRef\]](#)
23. Wei, J.-T.; Lin, S.-Y.; Wu, H.-H. A Review of the Application of RFM Model. *Afr. J. Bus. Manag. Dec. Spec. Rev.* **2010**, *4*, 4199–4206.
24. Djuricic, V.; Kascelan, L.; Rogic, S.; Melovic, B. Bank CRM Optimization Using Predictive Classification Based on the Support Vector Machine Method. *Appl. Artif. Intell.* **2020**, *34*, 941–955. [\[CrossRef\]](#)
25. Dogan, O.; Ayçin, E.; Bulut, Z. Customer Segmentation by Using RFM Model and Clustering Methods: A Case Study in the Retail Industry. *Int. J. Contemp. Econ. Adm. Sci.* **2018**, *8*, 1–19.
26. Amoozad Mahdiraji, H.; Tavana, M.; Mahdiani, P.; Abbasi Kamardi, A.A. A Multi-Attribute Data Mining Model for Rule Extraction and Service Operations Benchmarking. *Benchmarking Int. J.* **2022**, *29*, 456–495. [\[CrossRef\]](#)
27. Parvaneh, A.; Tarokh, M.; Abbasimehr, H. Combining Data Mining and Group Decision Making in Retailer Segmentation Based on LRFMP Variables. *Int. J. Ind. Eng. Prod. Res.* **2014**, *25*, 197–206.
28. Peker, S.; Kocyigit, A.; Eren, P.E. LRFMP Model for Customer Segmentation in the Grocery Retail Industry: A Case Study. *Mark. Intell. Plan.* **2017**, *35*, 544–559. [\[CrossRef\]](#)
29. Wei, J.T.; Lin, S.Y.; Yang, Y.Z.; Wu, H.H. The Application of Data Mining and RFM Model in Market Segmentation of a Veterinary Hospital. *J. Stat. Manag. Syst.* **2019**, *22*, 1049–1065. [\[CrossRef\]](#)
30. Akhondzadeh-Noughabi, E.; Albadvi, A. Mining the Dominant Patterns of Customer Shifts Between Segments by Using Top-k and Distinguishing Sequential Rules. *Manag. Decis.* **2015**, *53*, 1976–2003. [\[CrossRef\]](#)
31. Mosaddegh, A.; Albadvi, A.; Sepehri, M.M.; Teimourpour, B. Dynamics of Customer Segments: A Predictor of Customer Lifetime Value. *Expert Syst. Appl.* **2021**, *172*, 114606. [\[CrossRef\]](#)
32. Seret, A.; vanden Broucke, S.K.; Baesens, B.; Vanthienen, J. A Dynamic Understanding of Customer Behavior Processes Based on Clustering and Sequence Mining. *Expert Syst. Appl.* **2014**, *41*, 4648–4657. [\[CrossRef\]](#)
33. Xu, D.; Tian, Y. A Comprehensive Survey of Clustering Algorithms. *Ann. Data Sci.* **2015**, *2*, 165–193. [\[CrossRef\]](#)
34. Aghabozorgi, S.; Seyed Shirkhorshidi, A.; Ying Wah, T. Time-Series Clustering—A Decade Review. *Inf. Syst.* **2015**, *53*, 16–38. [\[CrossRef\]](#)
35. Kaufman, L.; Rousseeuw, P.J. *Finding Groups in Data: An Introduction to Cluster Analysis*; Wiley Online Library: Hoboken, NJ, USA, 1990.
36. Sakoe, H.; Chiba, S. A Dynamic Programming Approach to Continuous Speech Recognition. In Proceedings of the Seventh International Congress on Acoustics, Budapest, Hungary, 18–26 August 1971; Volume 3, pp. 65–69.
37. Sakoe, H.; Chiba, S. Dynamic Programming Algorithm Optimization for Spoken Word Recognition. *IEEE Trans. Acoust. Speech Signal Process.* **1978**, *26*, 43–49. [\[CrossRef\]](#)
38. Wang, X.; Smith, K.; Hyndman, R. Characteristic-Based Clustering for Time Series Data. *Data Min. Knowl. Discov.* **2006**, *13*, 335–364. [\[CrossRef\]](#)
39. Ezugwu, A.E.; Ikotun, A.M.; Oyelade, O.O.; Abualigah, L.; Agushaka, J.O.; Eke, C.I.; Akinyelu, A.A. A Comprehensive Survey of Clustering Algorithms: State-of-the-Art Machine Learning Applications, taxonomy, challenges, and future research prospects. *Eng. Appl. Artif. Intell.* **2022**, *110*, 104743. [\[CrossRef\]](#)
40. Ahmad, A.; Dey, L. A K-Means Clustering Algorithm for Mixed Numeric and Categorical Data. *Data Knowl. Eng.* **2007**, *63*, 503–527. [\[CrossRef\]](#)
41. MacQueen, J. Some Methods for Classification and Analysis of Multivariate Observations. Available online: https://digitalassets.lib.berkeley.edu/math/ucb/text/math_s5_v1_article-17.pdf (accessed on 22 April 2024).
42. Fayyad, U.; Reina, C.; Bradley, P.S. Initialization of Iterative Refinement Clustering Algorithms. In Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 27–31 August 1998; pp. 194–198.
43. Antunes, C.; Oliveira, A.L. Temporal data mining: An overview. In Proceedings of the KDD Workshop on Temporal Data Mining, San Francisco, CA, USA, 26–29 August 2001; pp. 1–13.
44. Bezdek, J.C. *Pattern Recognition with Fuzzy Objective Function Algorithms*; Kluwer Academic Publishers: London, UK, 1981.
45. Dunn, J.C. A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Cybern. Syst.* **1973**, *3*, 32–57. [\[CrossRef\]](#)
46. Krishnapuram, R.; Joshi, A.; Nasraoui, O.; Yi, L. Low-Complexity Fuzzy Relational Clustering Algorithms for Web Mining. *IEEE Trans. Fuzzy Syst.* **2001**, *9*, 595–607. [\[CrossRef\]](#)
47. Shavlik, J.W.; Dietterich, T.G. *Readings in Machine Learning*; Morgan Kaufmann: Cambridge, MA, USA, 1990.

48. Wang, X.; Smith, K.A.; Hyndman, R.J.; Alahakoon, D.A. Scalable Method for Time Series Clustering. Available online: <https://api.semanticscholar.org/CorpusID:8168184> (accessed on 21 April 2024).
49. Andreopoulos, B.; An, A.; Wang, X. A Roadmap of Clustering Algorithms: Finding a Match for a Biomedical Application. *Brief. Bioinform.* **2009**, *10*, 297–314. [[CrossRef](#)]
50. Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In Proceedings of the Knowledge Discovery and Data Mining, Portland, OR, SUA, 2–4 August 1996.
51. Chandrakala, S.; Chandra, C. A Density-Based Method for Multivariate Time Series Clustering in Kernel Feature Space. In Proceedings of the IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong, China, 1–8 June 2008; pp. 1885–1890.
52. Wang, W.; Yang, J.; Muntz, R. STING: A Statistical Information Grid Approach to Spatial Data Mining. In Proceedings of the International Conference on Very Large Data Bases, San Francisco, CA, USA, 25–29 August 1997; pp. 186–195.
53. Sheikholsami, G.; Chatterjee, S.; Zhang, A. WaveCluster: A Multi-Resolution Clustering Approach for Very Large Spatial Databases. In Proceedings of the International Conference on Very Large Data Bases, San Francisco, CA, USA, 24–27 August 1998; pp. 428–439.
54. Aghabozorgi, S.; Wah, T.Y.; Herawan, T.; Jalab, H.A.; Shaygan, M.A.; Jalali, A.A. Hybrid Algorithm for Clustering of Time Series Data Based on Affinity Search Technique. *Sci. World J.* **2014**, *2014*, 562194. [[CrossRef](#)] [[PubMed](#)]
55. Lai, C.-P.; Chung, P.-C.; Tseng, V.S. A Novel Two-Level Clustering Method for Time Series Data Analysis. *Expert Syst. Appl.* **2010**, *37*, 6319–6326. [[CrossRef](#)]
56. Venkataanusua, P.; Anuradha, C.; Murty, P.S.R.C.; Kiran, C.S. Detecting Outliers in High-Dimensional Datasets Using Z-Score Methodology. *Int. J. Innov. Technol. Explor. Eng. (IJITEE)* **2019**, *9*, 48–53. [[CrossRef](#)]
57. Chikodili, N.B.; Abdulmalik, M.D.; Abisoye, O.A.; Bashir, S.A. Outlier Detection in Multivariate Time Series Data Using a Fusion of K-Medoid, Standardized Euclidean Distance, and Z-Score. In *Information and Communication Technology and Applications. ICTA 2020. Communications in Computer and Information Science*; Misra, S., Muhammad-Bello, B., Eds.; Springer: Cham, Switzerland, 2021. [[CrossRef](#)]
58. Yıldız, E.; Güngör Şen, C.; Işık, E.E. A Hyper-Personalized Product Recommendation System Focused on Customer Segmentation: An Application in the Fashion Retail Industry. *J. Theor. Appl. Electron. Commer. Res.* **2023**, *18*, 571–596. [[CrossRef](#)]
59. Hughes, A.M. Boosting Response with RFM. *Mark. Tools* **1996**, *3*, 48.
60. Hosseini, S.M.S.; Maleki, A.; Gholamian, M.R. Cluster Analysis Using a Data Mining Approach to Develop CRM Methodology to Assess Customer Loyalty. *Expert Syst. Appl.* **2010**, *37*, 5259–5264. [[CrossRef](#)]
61. Wei, J.-T.; Lin, S.-Y.; Weng, C.-C.; Wu, H.-H. A Case Study of Applying LRFM Model in Market Segmentation of a Children's Dental Clinic. *Expert Syst. Appl.* **2012**, *39*, 5529–5533. [[CrossRef](#)]
62. Li, D.-C.; Dai, W.-L.; Tseng, W.-T. A Two-Stage Clustering Method to Analyze Customer Characteristics to Build Discriminative Customer Management: A Case of Textile Manufacturing Business. *Expert Syst. Appl.* **2011**, *38*, 7186–7191. [[CrossRef](#)]
63. Han, J.; Kamber, M.; Pei, J. *Data Mining: Concepts and Techniques*; Elsevier Science: Amsterdam, The Netherlands, 2011.
64. Paparrizos, J.; Gravano, L. k-Shape: Efficient and Accurate Clustering of Time Series. *ACM SIGMOD Rec.* **2016**, *45*, 69–76. [[CrossRef](#)]
65. Pearl, J. Theoretical Impediments to Machine Learning with Seven Sparks from the Causal Revolution. Technical report. In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, Marina Del Rey, CA, USA, 5–9 February 2018.
66. Pearl, J. *Causality: Models, Reasoning, and Inference*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2009.
67. Cheng, D.; Zhu, Q.; Huang, J.; Wu, Q.; Yang, L. A Novel Cluster Validity Index Based on Local Cores. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 985–999. [[CrossRef](#)]
68. Desgraupes, B. Clustering Indices. Available online: <https://cran.r-project.org/web/packages/clusterCrit/vignettes/clusterCrit.pdf> (accessed on 21 April 2024).
69. Paparrizos, J.; Gravano, L. Fast and Accurate Time-Series Clustering. *ACM Trans. Database Syst.* **2017**, *42*, 13. [[CrossRef](#)]
70. Ramos, P.; Santos, N.; Rebelo, R. Performance of State Space and ARIMA Models for Consumer Retail Sales Forecasting. *Robot. Comput. Integr. Manuf.* **2015**, *34*, 151–163. [[CrossRef](#)]
71. Chouakria, A.D.; Nagabhushan, P.N. Adaptive Dissimilarity Index for Measuring Time Series Proximity. *Adv. Data Anal. Classif.* **2007**, *1*, 5–21. [[CrossRef](#)]
72. Montero, P.; Vilar, J.A. TSclust: An R Package for Time Series Clustering. *J. Stat. Softw.* **2014**, *62*, 1–43. [[CrossRef](#)]
73. Batista, G.; Keogh, E.; Tataw, O.; Alves de Souza, V. CID: An Efficient Complexity-Invariant Distance for Time Series. *Data Min. Knowl. Discov.* **2013**, *28*, 634–669. [[CrossRef](#)]
74. Grandini, M.; Bagli, E.; Visani, G. Metrics for Multi-Class Classification: An Overview. *arXiv* **2020**, arXiv:2008.05756.
75. Martínez, F.; Charte, F.; Frias, M.P.; Martínez-Rodríguez, A.M. Strategies for Time Series Forecasting with Generalized Regression Neural Networks. *Neurocomputing* **2022**, *491*, 509–521. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.