IJSES

International Journal of Scientific Engineering and Science
Volume 9, Issue 2, pp. 102-105, 2025. ISSN (Online): 2456-7361

Data Segmentation Methods for Predicting Customer Behavior

Ansu Mathai Samuel
Senior Business Analyst at GoDaddy
Tempe, Arizona, USA

Abstract— *The article examines modern methods used in data segmentation for analyzing customer behavior across various sectors. The aim is to explore existing methodologies for dividing data and subsequently utilizing them to predict consumer preferences and behavior. The study also discusses clustering methods and the potential application of neural networks, which can identify hidden patterns, structure datasets, and improve prediction accuracy. The methodological foundation of the article encompasses algorithms for customer interaction and social preferences based on purchase data. Open-access scientific research demonstrates that analytical methods enhance prediction accuracy, thereby minimizing errors resulting from misinterpretation of initial data. The results confirm that data segmentation can predict consumer behavior and thus assist in developing precise marketing strategies and improving customer service. Therefore, the materials presented in this article will be useful for analysts, marketing specialists, and other researchers working with data. In summary, the findings clearly illustrate that selecting appropriate segmentation methods directly influences the accuracy of predictions regarding individual customer behavior and the ability to create tailored specific offers.*

Keywords— *Data segmentation, customer behavior prediction, machine learning, clustering, marketing strategies, data analytics.*

I. Introduction

Predicting customer demand enables companies to adjust their marketing strategies and minimize risks associated with fluctuating product and service demand. Data segmentation serves as a key tool for achieving this, allowing the identification of client groups with similar characteristics and facilitating behavioral prediction. Modern data processing methods and machine learning technologies offer the ability to analyze consumer preferences with reduced time costs, making this topic particularly relevant today.

Data segmentation supports the development of personalized strategies, improves customer service, enhances loyalty, and the choice of method depends on the specific task and conditions.

Scientific studies on data segmentation for predicting customer behavior examine various methodological combinations based on data types. These studies often explore the integration of traditional statistical techniques with modern algorithms to improve prediction accuracy. Within this article, a segmentation method that uses hybrid models—combining the advantages of multiple approaches—will be proposed.

The aim is to review existing data partitioning methodologies and their subsequent use in predicting consumer preferences and behaviors.

II. Materials and Methods

The article by Husein A. M. et al. [4] proposes a consumer clustering algorithm that focuses on customer preferences. In the work of Bartels C. [5], the use of open banking data for segmentation is described, enabling the creation of offers tailored to client interests. This approach helps identify groups with similar characteristics, thereby enhancing marketing effectiveness.

Akter S. et al. [3] present a methodology for machine data analysis aimed at predicting changes in consumer preferences within the digital marketing sphere. These technologies allow for the anticipation of customer needs at their early stages, enabling communication before explicit requests arise. This approach helps companies prepare for shifts in consumer behavior by adapting strategies in advance.

Abbasimehr H. and Shabani M. [7] discuss a time series clustering method applied in the banking sector. This approach accounts for temporal dynamics, resulting in more accurate forecasting. Articles by Rahim M. A. et al. [9] and Ernawati E., Baharin S. S. K., and Kasmin F. [6] examine the use of RFM analysis to predict repeat purchase likelihood, which helps classify customers and forecast their behavior. This method has proven effective in short-term forecasting in areas such as retail and financial services.

Zhou J., Wei J., and Xu B. [10] highlight how time series processing reveals patterns that are not accessible through static methods. The work of Abbasimehr H. and Shabani M. [1] demonstrates that considering temporal dynamics enables the creation of accurate forecasts that reflect long-term changes in client behavior.

Pratheebha T. et al. [2] present a model for predicting customer churn through optimization methods and individualized approaches. In other fields, such as financial services and insurance, these technologies help identify consumer preferences and formulate offers aligned with client demands, as discussed in Nimmagadda V. S. P. [8].

Thus, analytical methods allow for accurate forecasting and the development of personalized strategies, which enhance marketing effectiveness and improve business processes.

The practical section of this study draws upon sources [11-14]. Source [11], hosted on Amazon's official website, describes how the company employs segmentation methods in its operations. Source [12], found on www.almabetter.com, presents the experience of Netflix. Source [13], hosted on diversedaily.com, details Spotify’s approach. The final source, [14], published on Airbnb’s official website, provides information on how the company incorporates segmentation in its activities.

Clustering methods and RFM analysis are used to segment clients, allowing for the identification of groups with similar behavioral characteristics and predicting their actions shortly. Modern analytical approaches that account for a large number of factors are well-suited for long-term forecasts. These technologies enhance predictive accuracy, creating new planning opportunities. However, the issue of transparency in analytical model operations remains relevant.

Published academic works show that the application of analytical methods improves prediction accuracy and reduces the likelihood of errors caused by incorrect data interpretation.

III. Results and Discussion

Data segmentation as a method of analysis remains a key tool for predicting consumer behavior in a rapidly changing environment where data originates from diverse sources. This approach is crucial for extracting information from large and often unstructured datasets. Segmentation identifies consumer groups and improves prediction accuracy regarding their future actions, which is essential for marketing, loyalty, and retention strategies. Methods of data segmentation for forecasting consumer behavior involve several approaches, each with distinct characteristics, limitations, and suitability based on data type, analytical goals, and available computational resources [4, 5]. Figure 1 below presents data segmentation methods.

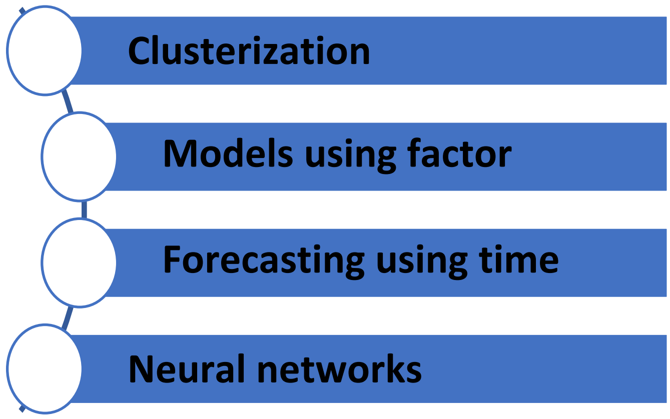

Figure 1. Data segmentation methods [4, 5, 7]

Clustering remains an important technique for identifying homogeneous groups, and modern methods enhance the capabilities of classical algorithms, enabling more precise segmentation. The weighted k-means method, for example, considers differences in feature importance when calculating cluster centroids, thus influencing the outcome.

Hierarchical clustering does not require a predefined number of groups, allowing the construction of dendrograms that reveal data structures and consumer preference patterns, which are essential for designing marketing strategies.

Factor analysis helps identify latent variables influencing interrelations among characteristics. In predicting customer behavior, this method detects factors such as a propensity toward specific consumption types or sensitivity to price changes. Principal component analysis (PCA) and independent component analysis (ICA) reduce data dimensionality, highlighting key features for segmentation—critical when dealing with large datasets.

Time-series analysis is employed for segmentation since consumer behavior evolves over time. ARIMA models and recurrent neural networks consider temporal dependencies, aiding in the prediction of purchase cycles and responses to marketing initiatives. Time-series models can be combined with clustering to analyze the dynamics of different segments, such as identifying groups with varying loyalty levels that respond differently to seasonal discounts or price fluctuations [1, 3, 7].

Another method for analyzing customer behavior is RFM analysis, which considers three parameters: the most recent purchase (Recency), the frequency of purchases (Frequency), and the monetary value of customer spending (Monetary). High values across all parameters indicate loyal customers who should be retained through personalized offers [2, 6, 8].

Neural networks are capable of identifying relationships by considering both explicit and hidden parameters of customer behavior. For example, convolutional neural networks (CNNs) are used to analyze images associated with user preferences, while recurrent neural networks (RNNs) analyze human behavior based on sequences of actions, such as purchases and product views, enabling the processing of large datasets. The existing segmentation stages are illustrated for clarity in Figure 2.

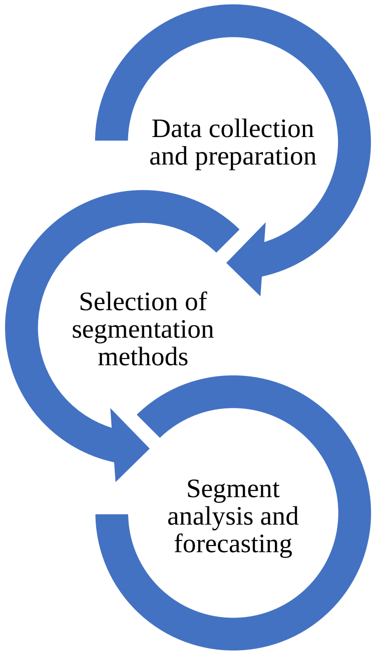

Figure 2. Segmentation stages

Before segmentation, it is necessary to collect and prepare data. Sources include transaction information with parameters such as date, amount, product, and payment method; behavioral data that captures user actions on a website or in an app; and demographic data that identifies age, gender, income, location, and marital status.

IJSES

International Journal of Scientific Engineering and Science

Volume 9, Issue 2, pp. 102-105, 2025. ISSN (Online): 2456-7361

Customer behavior is predicted using machine learning models, examples of which include:

- Logistic regression, which analyzes the likelihood of a purchase based on the user's prior behavior.
- Random Forest, which employs decision trees to process large datasets.
- XGBoost is used for handling complex dependencies, such as predicting customer churn.
- Neural networks, which identify patterns and forecast behavior by analyzing interrelated data.

After segmentation is complete, each customer group's behavior is analyzed. Understanding these factors helps more accurately plan future actions. Subsequently, segmentation is used as a basis for developing strategies to engage with customers [3, 7, 9, 10].

The subsequent analysis involves a comparison of data segmentation methods, which is presented in Table 1.

TABLE 1. Comparison of data segmentation methods

Method Name	Description	Advantages	Disadvantages	Use Cases	Impact on Forecasting
Segmentation by Demographic Factors	Customers are divided based on age, gender, income level, residence, and other demographic attributes.	Easy data collection and subsequent analysis; facilitates rapid customer classification.	Does not always reflect customer preferences.	Marketing, targeted advertising, identifying potential purchasing behavior.	Can aid in classification but is less precise than other methods for predictions.
Segmentation by Behavioral Traits	Analyzes customer behavior.	Helps predict preferences and needs, improving the personalization of offers.	Requires data collection and analysis; interpreting behavioral patterns can be complex.	Personalized recommendations, improving customer experience, predicting purchases.	Enhances forecasting accuracy by identifying hidden behavioral patterns.
Segmentation Using Machine Learning	Employs machine learning algorithms to uncover hidden patterns and segments in data.	Highly accurate; adapts to changes over time.	Requires computational resources and expertise to configure models.	Churn prediction, enhancing recommendations, dynamic offer adjustments.	Improves prediction accuracy and helps adapt to change.
Segmentation Based on Customer Lifecycle	Customers are categorized by their lifecycle stage.	Helps understand consumer needs at different journey stages.	Challenging to implement; requires regular data updates and tracking customer behavior.	Customer retention, marketing campaigns, and enhancing customer service.	Impacts retention, and repeat sales, and assists in forecasting customer needs.
Geographic Segmentation	Divides customers by regions, cities, countries, or even neighborhoods within a city.	Easy to analyze and implement; useful for marketing strategies.	Lower accuracy when customer behavior is not heavily influenced by location.	Localized marketing strategies, improving delivery, logistics, and regional targeting.	Helps forecast needs in specific geographic areas, but with limited accuracy.

Next, let us consider the experience of certain companies. For example, Amazon employs behavior-based data segmentation. Machine learning technologies and recommendation systems are used to analyze purchasing activity and preferences. Each user is categorized according to various criteria, including purchase history and product interactions. Algorithms assist in generating recommendations that predict which products may interest the user [11].

Netflix implements viewer-oriented segmentation, analyzing content preferences through machine learning. The company develops recommendations for movies and series by considering individual tastes and the preferences of other viewers with similar interests. Analysis of user satisfaction with content allows the prediction of new suggestions, minimizing churn. This approach increases engagement and enhances the quality of recommendations [12].

Spotify utilizes data on musical preferences to curate personalized playlists, such as Discover Weekly and Release Radar. Algorithms analyze user tastes to predict new musical directions that may capture their interest. Preference-based segmentation helps identify artists and genres aligned with user interests, thereby increasing platform usage time and reducing churn [13].

Airbnb applies segmentation based on user preferences and booking data analysis. Machine learning algorithms recommend accommodations tailored to user interests, predict popular destinations, and optimize listings. This approach improves conversion rates and facilitates price adjustments [14].

Thus, the task of data segmentation for predicting client behavior calls for a combination of traditional methods and machine learning technologies. Effectively solving these challenges relies on a strong theoretical foundation, practical experience with data, and the application of modern technological solutions.

IV. Conclusion

The analysis of various approaches has demonstrated their ability to uncover patterns and predict consumer preferences. Hybrid models that combine statistical methods with modern algorithms produce more accurate results.

The findings confirm that the choice of segmentation method depends on the characteristics of the data and the specifics of the business. Accounting for these factors increases forecasting accuracy and reduces the likelihood of errors in decision-making. The article contributes to the advancement of consumer behavior forecasting by proposing new approaches to leveraging analytical technologies in business.

Volume 9, Issue 2, pp. 102-105, 2025. *ISSN (Online): 2456-7361*

REFERENCES

1. 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 //Journal of ambient intelligence and humanized computing. – 2021. – Vol. 12 (1). – pp. 515-531.
2. Pratheebha T., Megala S. S., Indumathi V. An Improved Churn Prediction Model using Relevance Feature Identification and Personalized Modeling Technique with Optimized Retention Strategies //2024 8th International Conference on Electronics, Communication and Aerospace Technology (ICECA). – IEEE, 2024. – pp. 1489-1496.
3. Akter S. et al. The effects of artificial intelligence and machine learning in consumer behavior, demand prediction, and chatbots in the realm of digital marketing // Business Forecasting and Marketing Intelligence. - 2024. – pp. 1-25.
4. Husein A. M. et al. Clustering algorithm for determining marketing targets based customer purchase patterns and behaviors //Sinkron: jurnal dan penelitian teknik informatika. – 2021. – Vol. 5 (2B). – pp. 137-143.
5. Bartels C. Cluster analysis for customer segmentation with open banking data //2022 3rd Asia Service Sciences and Software Engineering Conference. – 2022. – pp. 87-94.
6. Ernawati E., Baharin S. S. K., Kasmin F. A review of data mining methods in RFM-based customer segmentation //Journal of Physics: Conference Series. – IOP Publishing, 2021. – Vol. 1869 (1) . – pp. 1-9.
7. Abbasimehr H., Shabani M. A new methodology for customer behavior analysis using time series clustering: A case study on a bank's customers //Kybernetes. – 2021. – Vol. 50 (2). – pp. 221-242.
8. Nimmagadda V. S. P. Artificial Intelligence for Customer Behavior Analysis in Insurance: Advanced Models, Techniques, and Real-World Applications //Journal of AI in Healthcare and Medicine. – 2022. – Vol. 2 (1). – pp. 227-263.
9. Rahim M. A. et al. RFM-based repurchase behavior for customer classification and segmentation //Journal of Retailing and Consumer Services. – 2021. – Vol. 61. – pp. 102566.
10. Zhou J., Wei J., Xu B. Customer segmentation by web content mining //Journal of Retailing and Consumer Services. – 2021. – Vol. 61. – pp. 102588.
11. Amazon Segmentation, Targeting and Positioning: Widest Range of Target Customer Segment. [Electronic resource] Access mode: https://research-methodology.net/amazon-segmentation-targeting-positioning-widest-range-target-customer-segment-2 / (date of access: 12.01.2025).
12. How Netflix Uses ML & AI For Better Recommendation for Users. [Electronic resource] Access mode: https://www.almabetter.com/bytes/articles/how-netflix-uses-machine-learning-to-keep-users-glued-to-the-screen (date of request: 12.01.2025).
13. Spotify: Analyzing Personalized Playlists, Podcast Offerings, and Artist Discovery Features. [Electronic resource] Access mode: https://diversedaily.com/spotify-analyzing-personalized-playlists-podcast-offerings-and-artist-discovery-features / (date of access: 12.01.2025).
14. Airbnb Segmentation, Targeting & Positioning. [Electronic resource] Access mode: https://research-methodology.net/airbnb-segmentation-targeting-positioning / (date of access: 12.01.2025).