Volume 3, Issue 5, May - 2025 **ISSN (E):** 2938-3811

Web of Scientists and Scholars: Journal of Multidisciplinary Research

APPLICATION OF DATA ANALYSIS FOR PREDICTING CONSUMER BEHAVIOR

Aizhan Adilkhan
Master's Degree in Technology Management,
University of California Santa Barbara, USA

Abstract
This article explores the importance of analyzing consumer behavior in the context of the digital economy and increasing market competition. It highlights the role of data analysis and modern Data Science technologies in shaping personalized marketing strategies, predicting customer behavior, and making well-informed managerial decisions. Special attention is given to the practical application of analytical methods such as clustering, association rule analysis, logistic regression, decision trees, machine learning methods, and time series analysis. The article provides successful examples of how these approaches are used by major international companies such as Spotify, Walmart, Netflix, Uber, and others. The advantages and limitations of each method are outlined, enabling a reasoned selection of analytical tools to solve specific business challenges.

Keywords: data analysis, consumer behavior, machine learning, forecasting, personalization, marketing.

Introduction

Scientific Novelty

This article proposes a comprehensive methodology for applying modern analytical methods of big data processing to predict consumer behavior in the context of the digital economy. Unlike traditional models such as AIDA, which do not account for the individuality and dynamic nature of consumer choices, the developed approach integrates semantic and statistical analysis using machine learning algorithms, including clustering methods, association rule mining, logistic regression, decision trees, and ensembles (Random Forest, XGBoost).

For the first time, the effectiveness of hybrid analytical tools based on unstructured data from open sources and social networks is demonstrated to create personalized marketing strategies and improve the accuracy of customer behavior prediction. The article also identifies the advantages and limitations of various methods with practical examples from leading companies in retail, finance, and digital services, providing a rationale for selecting optimal tools for specific business tasks.

Additionally, the potential of time series data analysis is revealed to account for seasonal and trend fluctuations in consumer behavior, significantly enhancing the quality of strategic planning and inventory management. The presented work contributes to the development of Data Science theory and practice in marketing and justifies the need for an integrated approach combining technical, analytical, and ethical aspects in predicting consumer behavior in modern conditions.

In today's world, organizations focus not only on meeting current customer needs but also on predicting their future demands. This is driven by intense market competition, in which forward-thinking companies strive not only to produce high-quality products but also to achieve maximum personalization by taking into account the needs and preferences of individual customers. It should be noted that full individualization of offerings remains a complex task for production processes. The reliability and completeness of information obtained about consumers has a significant impact on the effectiveness of a company's performance.

At present, a pressing issue is the so-called *"blurring"* of the core target audience — a mismatch between actual consumers and their presumed characteristics. Marketers possess vast datasets, yet they are not always able to accurately identify which potential or existing customers are most likely to make a purchase.

In this context, analyzing consumer purchasing behavior becomes especially important for predicting future actions. Advances in data collection and processing technologies, along with the availability of high-performance computing systems, have significantly improved the quality of marketing analytics. Data-driven methodologies have become a key component of strategic planning in sectors such as retail, banking, telecommunications, and other areas of the economy.

Artificial intelligence provides businesses with the ability to process vast volumes of information — including customer online activity, purchasing patterns, and behavioral tendencies — and to promptly influence decision-making processes. Understanding consumer behavior mechanisms has shifted from being a desirable practice to becoming a fundamental requirement for the successful operation of any organization. Every user action in the digital environment — whether interacting with a website or making a purchase — offers valuable insights into their preferences and needs.

Consumer behavior encompasses a range of actions and processes related to the selection, purchase, use, and disposal of goods and services, as well as the factors influencing these behaviors. According to the well-known economist and marketer Philip Kotler, consumer behavior includes the actions of individuals involved in obtaining and using products, along with the decision-making processes preceding these actions [1]. Similarly, researchers L.G. Schiffman and L.L. Kanuk define this term as the process by which individuals, groups, and organizations select, buy, use, and dispose of goods and services to satisfy their needs and desires [2].

An important aspect of consumer behavior involves motivation, perception, and value assessment, which precede and accompany the purchase process. Henry Assael points out that consumer behavior encompasses not only the act of purchasing itself but also the preceding motivational stages and subsequent behavior, including satisfaction with the acquired product [3]. G.A. Churchill and J. Peter emphasize the influence of both internal and external factors — such as perception, learning, personality traits, culture, and social environment — in shaping consumer behavior [4].

Russian scholar Professor I.V. Romanenko defines consumer behavior as a system of socio-economic and psychological characteristics of an individual as manifested in their market activity [5].

Thus, consumer behavior is a process of interaction between a company and its customers. Throughout this process, consumers make choices and purchase goods and services. The use of analytical methods in managing consumer behavior allows organizations to gain additional competitive advantages and obtain objective results. Each year, the volume of consumer data grows, enabling more effective analysis of user activity.

Modern predictive analytics systems, based on decision-making informed by past events, can significantly enhance the consumer's interaction with a product. By analyzing data, organizations can unlock the following opportunities [8]:

Capabilities	Description of Capabilities
Segmentation and Categorization of the Customer Base	Efficient allocation of new customers to existing consumer groups or the creation of entirely new categories for their classification.
Comprehensive Customer Analysis ("Customer 360")	Creation of a comprehensive customer profile aimed at an in-depth interpretation of their behavior and consumer priorities.
Personally tailored advertising campaigns and special discounts.	Development of offers tailored to the individual preferences of the customer. Promotions targeted at specific consumers.
Monitoring and responding to sudden increases in consumer activity.	Monitoring peak loads in consumer behavior and responding promptly to them.

(Table 1 – Data Analysis Capabilities for Organizations)

Traditional concepts of consumer behavior, such as the AIDA model (Attention - Interest - Desire - Action), offer a basic algorithm but do not take into account the uniqueness of each consumer and the variability of their choices in the modern digital era. In this context, analytical data processing methods play a key role.

Currently, automated planning and decision support systems actively use methods of processing and analyzing data obtained from open sources. For example, suppliers of goods and services across various industries integrate social media user behavior analysis into the development of their marketing strategies.

Developing such software solutions requires processing large volumes of data through parallel computing. The application of semantic and statistical analysis of semi-structured information enables the creation of highly efficient analytical tools [6].

Raw data extracted from sources is often unprocessed; however, the use of analytical approaches allows for the interpretation of these complex data sets, which helps identify key causal relationships in observed trends and develop strategies to extract commercial benefits from the obtained information [7].

To accurately predict consumer behavior, which plays an important role in strategic planning and business process improvement, a variety of data analysis methods are applied. These

methods help detect patterns, create models, and forecast future customer actions based on the analysis of past and current data.

It is worth noting that when forecasting consumer behavior, the choice of analytical method, as mentioned above, plays a crucial role in ensuring the accuracy and interpretability of the results. Each method has a unique set of advantages and limitations that must be considered when applying it.

Clustering methods, such as K-means and hierarchical clustering, demonstrate effectiveness in targeting marketing campaigns by dividing the customer base into homogeneous segments, followed by personalized offers. The simplicity of implementation and interpretation (especially for K-means) is an additional advantage. However, clustering is sensitive to the choice of initial parameters, and interpreting the results of hierarchical clustering becomes more complex as data volume grows. Also, these methods may fail to detect complex nonlinear dependencies.

The popular company Spotify uses the K-means algorithm to segment users by musical tastes, enabling personalized playlists and recommendations. As a result, average listening time increases, helping to retain the audience. In the retail sector, the perfume and cosmetics chain Sephora employs hierarchical clustering to group customers based on behavior and preferences, designing targeted promotions and loyalty programs [9].

Thus, despite difficulties in interpreting results when working with large datasets, this method contributes to increased customer engagement and sales growth.

Association rule analysis (market basket analysis) provides valuable insights for developing recommendation systems, optimizing product placement, and conducting cross-selling. At the same time, the number of generated rules can be excessive, and processing large volumes of data requires significant computational resources.

A typical example is the American company Walmart, which applies this approach to study combinations of products frequently purchased together. The resulting data help improve product placement and cross-selling, contributing to an increase in the average ticket size. Similarly, companies such as Amazon, Ozon, Wildberries, and AliExpress use these algorithms to create “Frequently Bought Together” recommendations, significantly improving suggestion accuracy.

Logistic regression and decision trees, as classification methods, provide interpretable results that help understand factors influencing the target event (e.g., making a purchase or customer churn). Ease of implementation and interpretation are also advantages. However, logistic regression assumes a linear relationship between variables, limiting its applicability in complex situations, while decision trees are prone to overfitting on small datasets.

In banking, one of the largest global banks, JPMorgan Chase, uses logistic regression to assess client credit risk, helping to reduce losses from loan defaults. Similarly, the financial company American Express employs decision trees to identify customers at high risk of service cancellation and develop retention strategies. Thanks to the simple interpretability of these models, business analysts can quickly identify key factors and make informed decisions.

More advanced machine learning methods, such as Random Forest, XGBoost, and neural networks, provide high accuracy in predicting complex behavior models when large amounts of historical data are available. Their ability to model nonlinear dependencies is a significant advantage. However, these methods require substantial computational resources, are complex to interpret ("black box"), and can be prone to overfitting if not properly tuned.

The entertainment company and streaming service Netflix successfully uses advanced machine learning methods, including ensembles of decision trees and neural networks, to predict user preferences and create personalized content recommendations. This improves engagement and reduces subscriber churn. In finance, one of the world's largest investment banks, Goldman Sachs, applies XGBoost to develop scoring models and forecast market trends for better-informed investment decisions. Despite the high computational costs and interpretability challenges, the accuracy of these approaches justifies their use.

Time series analysis is effective in predicting behavior based on temporal dynamics, considering trends and seasonal fluctuations. It accounts for temporal dependence between observations. However, it requires sufficient historical data, is sensitive to outliers and anomalies, and assumes stationarity or the ability to transform the time series into a stationary one.

The international public company Uber, a taxi, delivery, and related services aggregator, widely uses time series approaches to estimate service demand across regions and times of day [10]. This improves driver allocation and reduces customer waiting times. In retail, the American company operating the Target and SuperTarget store chains uses time series analysis to forecast seasonal demand for various product categories, aiding effective inventory planning and cost reduction.

Thus, selecting a specific data analysis method for predicting consumer behavior requires careful consideration of the task, available data, and acceptable complexity. Proper use of these methods enables companies to make informed decisions, improve marketing approaches, and strengthen their competitive advantage.

Predicting consumer behavior using data analysis is becoming an important part of competitive strategies in the modern economy. Innovative Data Science technologies help identify patterns in customer actions and effectively apply them for personalization, retention, and expansion of the customer base. Successful implementation of such methods requires combining technical tools, analytical approaches, and adherence to ethical standards.

References

1. Kotler F., Keller K. L. Marketing Management: Textbook / Translated from English – 12th ed. – St. Petersburg: Piter, 2007. – 816 p.
2. Schiffman L. G., Kanuk L. L. Consumer Behavior: Textbook / Translated from English – Moscow: Williams, 2003. – 784 p.
3. Assel G. Consumer Behavior: Study Guide / Translated from English – Moscow: INFRA-M, 1999. – 971 p.

4. Churchill G. A., Peter J. P. Marketing: Creating Consumer Value / Translated from English – St. Petersburg: Piter, 2005. – 624 p.
5. Romanenko I. V. Consumer Behavior: Theory and Practice: Study Guide. – Moscow: Economics, 2010. – 224 p.
6. Ivashchenko A. V., Stolbova A. A. Recursive Decomposition in Big Data Market: Example of Social Media User Purchasing Activity Analysis // Models, Systems, Networks in Economics, Technology, Nature, and Society. – 2019. – No. 4(32). – pp. 5–15.
7. Ivanchenko O. V. Data Mining and Business Analytics in Business Management and Marketing // Bulletin of Rostov State Economic University (RINH). – 2022. – No. 4(80). – pp. 125–130.
8. Loktionova E. A., Ragozina A. V. Features of Big Data Analysis Systems in Commercial Banking // Baikal Research Journal. – 2017. – Vol. 8, No. 2. – p. 9.
9. Sephora Marketing Strategy // Business Model Analyst: [website]. – URL: https://businessmodelanalyst.com/ru/%D0%BC%D0%B0%D1%80%D0%BA%D0%B5%D1%82%D0%B8%D0%BD%D0%B3%D0%BE%D0%B2%D0%B0%D1%8F-%D1%81%D1%82%D1%80%D0%B0%D1%82%D0%B5%D0%B3%D0%B8%D1%8F-sephora/ (accessed 16.05.2025).
10. How Walmart Uses Big Data // Habr: [website]. – URL: https://habr.com/ru/articles/457462/(accessed16.05.2025).

Web of Scientists and Scholars: Journal of Multidisciplinary Research