

Review on the Use of Data Analysis for Customer Segmentation and Personalization in Marketing Strategies

Mengzhen Xiao^{1,a,*}, Yongchao Xu^{2,b}, Zekai Gao^{3,c}

¹*Wuhan Britain-China School, Wuhan, Hubei Province, 430030, China*

²*International Department, Zhenhai High School, Ningbo, Zhejiang Province, 315100, China*

³*Jiangsu Taizhou High School, Taizhou, Jiangsu Province, 225399, China*

a. 1195681043@qq.com, b. xyongchao2023@163.com, c. zekaigao@iCloud.com

**corresponding author*

Abstract: This paper studies the application of data analysis in the field of customer segmentation and personalization. With the advent of the era of big data, enterprises and organizations have rich data resources that contain important information about customer behavior, preferences, and purchasing habits. Through data analytics, companies can better understand customer needs and provide personalized products and services, thereby improving customer satisfaction and loyalty. This study aims to explore the application of data analytics in the field of customer segmentation and personalization, and provides some common data analysis methods. Data analysis methods include probabilistic methods and Bayesian methods, among others. Through case studies, the application of data analysis in the field of e-commerce is explained. The research results show that data analytics has important value and potential applications in customer segmentation and personalized marketing. However, data analysis still faces some challenges and limitations in practice, including data security and quality issues. Future research can continue to explore methods and techniques for data analysis, solve problems such as data privacy and security, and apply data analysis in more fields.

Keywords: data analysis, customer segmentation, personalized marketing, e-commerce.

1. Introduction

Fast-moving advancements in hardware, software, and communication technologies have aided in the development of web-connected sensory devices that collect data about the real world and report its observations. Between 25 and 50 billion Internet-connected gadgets are anticipated to be in use worldwide by the year 2020. The amount of data published will rise as these figures rise and technology advances [1]. The usefulness of using current data for study is rising in popularity at this time when researchers from all over the world are gathering and archiving enormous amounts of data [2]. In statistical data analysis, it is frequently beneficial to study and present a body of data as though it may have started out as a one-dimensional sample, that is, data that one intends to treat as an unstructured array for analytical purposes. This is applicable to "derived" data, such as the residuals from a model fitted to the data, more frequently than it is to "original" data [3].

In today's digital age, businesses and organizations have vast data resources that contain important information about customer behavior, preferences, and purchasing habits. In order to better

understand customer needs and provide personalized products and services, data analysis has become an important tool. Customer segmentation and personalized marketing have become one of the key factors of enterprise competitiveness. Through data analysis technology, enterprises can divide customers into different markets and carry out targeted marketing activities according to the needs of different markets. Therefore, it is of great theoretical and practical significance to study the application of data analysis in the field of customer segmentation and personalization.

This situation motivated us to take some research about the way that data analysis can be used in consumer segmentation and consider some methods of data analysis and explain them. Firstly, we will review the definition and application for personalization and customer segmentation. In addition, we will also explain the relationship between personalization and customer segmentation. Secondly, we will review some methods, such as probability methods and Bayesian methods for the data analysis and add some explanation on them. We will also use some case analysis in E-commerce and tourism to help to have a better understanding on the application of data analysis in customer segmentation and personalization.

2. Customer Segmentation and Personalization

2.1. Customer Segmentation

Customer segmentation is a technique that helps businesses better understand the needs and preferences of various customer groups and offer more targeted and relevant products, services, and marketing initiatives to draw in more customers. It divides a wide range of customer groups into similar characteristics, needs, or behaviors, etc. The degree of within-group homogeneity that distinguishes different customer segments makes it more likely that the members of a segment will react to marketing initiatives similarly. This enables businesses to allocate marketing resources to each category more effectively [4]. Through customer segmentation, enterprises can more accurately locate the market and demand of target customers and develop corresponding marketing strategies, the advantage of this is that it can improve the efficiency and effect of marketing, and increase customer loyalty, thereby bringing more sales opportunities and profits, but also can avoid the waste of resources for irrelevant people, improve the efficiency of advertising and some promotional resources.

In today's fiercely competitive corporate environment, customer relationship management (CRM) has emerged as a prominent business approach. CRM is a tool that uses Web/Apps technology to give businesses the capacity to comprehend their customers' or potential customers' typical behaviors and deliver specific activities that can persuade them to make decisions and transactions [5]. Companies are becoming more and more aware of the numerous potential advantages that CRM may offer. The following are a few possible advantages of CRM: (1) Enhanced customer loyalty and retention, (2) Increased customer profitability, (3) Customer value creation, (4) Customized products and services, and (5) Products and services with fewer steps and greater quality. Since customer satisfaction is given more weight in today's business environment, many companies are concentrating on the idea of customer loyalty and profitability in order to grow market share and improve customer satisfaction. CRM, the fundamental business idea to improve customer relationships, is becoming a firm's key competency. Understanding a customer's genuine worth and loyalty is the first step in developing a successful CRM for a business because this information can be used to launch more focused and individualized marketing campaigns [6].

2.2. Personalization

As long as human society has existed, people have felt the urge to customize their surroundings, equipment, and products to address their own issues. Personalization is now more feasible than ever

thanks to technical advancements, the Internet, and new media on a larger scale and with greater efficiency [7]. Designing technology and application features for the end-user experience involves using a toolkit called personalization.

Personalization has diverse conceptual meanings to different persons in different disciplines. Different conceptualizations, in turn, require various research implementations and procedures. While e-commerce marketers rely on user profiles and purchase information to segment customers, cognitive scientists utilize explicit mental modeling to distinguish users [7].

Personalized marketing is based on the individual differences and specific needs of each customer, providing each customer with a unique and customized experience and service. Through personalized marketing, companies can better meet the personalized needs of customers and make accurate recommendations and customization according to their preferences and behaviors. This personalized interaction can significantly increase customer engagement and satisfaction, cultivate customer loyalty, and win more word-of-mouth publicity and repeat purchases. The so-called customer is God, only through personalized marketing and adhering to the principle of customer first, constantly more perfect than other companies to meet customer needs, fit the taste of customers, in order to have a steady stream of customers and develop opportunities in this increasingly competitive market.

Personalization has diverse conceptual meanings to different persons in different disciplines. Different conceptualizations, in turn, require various research implementations and procedures. While e-commerce marketers rely on user profiles and purchase information to segment customers, cognitive scientists utilize explicit mental modeling to distinguish users [7]. Relevant concepts include personalization (or how much the system adapts the material to a user's preferences) and customization (where a user purposefully adapts the content by selecting options and/or coming up with new content) [8].

The implementation of only a few, if any, predefined personalization strategies is supported by the majority of the personalization systems currently in use. Personalization strategies are used to tailor learning scenarios to the learner profiles. Additionally, educators and researchers might pinpoint additional requirements for putting new personalization techniques into practice. There is no one personalization method that can be utilized to tailor learning scenarios, and the personalization parameters that are applied heavily influence each personalization method. The personalization of learning situations must be handled in accordance with the personalization strategies suggested by the person in charge of personalization when there are various personalization needs identified by educators, professors, and students [9].

3. Methods of Data Analysis

The application of data analytics in the field of customer segmentation and personalization is very important. Through data analytics, businesses can better understand customer needs and behaviors, leading to more personalized products and services, and increased customer satisfaction and loyalty. The following are some common data analysis methods and corresponding application cases.

3.1. Probability Methods

Probabilistic methods are a set of mathematical techniques used to analyze and quantify uncertainty and randomness. A framework for comprehending and calculating the possibility of various outcomes or events occurring is provided by probability theory.

The concept of probability, which quantifies the possibility that a specific event will occur, serves as the foundation for the probabilistic method. A number between 0 and 1, where 0 denotes improbability and 1 denotes certainty, is typically used to express probability. In any given event space, the total probability of all outcomes is always equal to 1 [3].

For one-dimensional samples, the empirical cumulative distribution function (e.c.d.f.), i.e. a plot of the i th ordered value as ordinate against $(t - \lfloor jn \rfloor)$ as abscissa, provides an exhaustive representation of the data, under the following broad assumptions: (i) that the order of the observations is immaterial; (ii) that there is no classification of the observations based on extraneous considerations which one wishes to employ; and (iii) if the sample is non-random, then appropriate weights are specified [3].

Probabilistic methods can be used to analyze and predict in a variety of fields, including statistics, finance, engineering, and the natural sciences. It is especially useful when dealing with uncertain or random phenomena, i.e., where the outcome cannot be determined.

There are different types of probabilistic methods, including:

1. Classical probability: This method assumes that all outcomes in the event space are equally likely. Often used when dealing with simple and clearly defined situations [3].
2. Conditional probability: This method calculates the probability of an event occurring if another event has already occurred. It is usually used when events are interdependent [3].
3. Bayesian probability: This method incorporates previous beliefs or knowledge into the analysis and updates the probability as new information becomes available. It is widely used in machine learning, data analysis and decision making. [3].

Probability distributions and random variables: Random variables are those whose values are decided by chance. The probability that various values of a given random variable will occur is described by a probability distribution. Normal distribution, binomial distribution, and Poisson distribution are examples of common probability distributions [3].

Probabilistic methods provide a rigorous and systematic way to analyze uncertainty and randomness. By quantifying probabilities, these methods help us make informed decisions, assess risk, and analyze data in a variety of applications.

3.2. Bayesian Methods

Bayesian methods are a set of statistical techniques that allow us to update our beliefs or probabilities about hypotheses or parameters as new evidence emerges. The key concept of the Bayesian approach is the use of Bayes' theorem, which relates the conditional probability of a given hypothesis to the prior probability of the hypothesis. In Bayesian methods, we start with an initial belief or prior probability distribution for a hypothesis or parameter. As new data is observed, we update the belief using Bayes' theorem, calculating a posterior probability distribution that represents the updated belief given by the observed data. This iterative process of updating beliefs with new data can continue as more data becomes available, resulting in more accurate and refined estimates [10].

Thomas Bayes, who developed the Bayes' rule, a straightforward mathematical relationship between conditional probabilities, gave the field of Bayesian analysis its name. The following is a possible way to express the rule when it applies to both parameters and data:

$$p(\theta|D) = p(D|\theta) p(\theta)/p(D)$$

where D stands for the observed data and θ refers to a vector of descriptive model parameters. The posterior distribution, defined as $p(\theta|D)$, indicates the relative plausibility of each parameter combination given the data. The distribution of credibility adds to 1 and is equivalent to a probability distribution because the parameter range defines the entire space of potential descriptions [11].

Bayesian methods have several advantages over classical statistical methods. First, it provides a consistent framework for incorporating prior knowledge or beliefs into the analysis, which is particularly useful when working with small or limited data sets. Second, Bayesian methods allow uncertainty to be quantified, providing a probability distribution rather than a single point estimate. This allows us to make more informed decisions and interpret the results in a probabilistic manner [11].

Bayesian methods have applications in fields as diverse as data analysis, machine learning, decision making, and medical research. It is especially useful when data is limited or when making predictions about future events. However, when working with complicated models or vast amounts of data, Bayesian approaches can be computationally demanding and necessitate precise prior distribution selection [11].

There are also some other methods such as:

1. Cluster analysis: Through cluster analysis, customers can be divided into different segments, and corresponding marketing strategies can be formulated according to the characteristics and behaviors of the group. Cluster analysis groups customers through similarity measures based on their characteristics and behavior data.

2. Association rule analysis: Through association rule analysis, you can discover the association relationship and purchase pattern between customers. For example, customers often buy product B at the same time as product A, which can provide a basis for cross-selling and recommendations for businesses.

3. Predictive analytics: Through predictive analytics, you can predict the future behavior of customers, such as purchase probability, churn risk, etc. This helps companies take targeted actions and develop appropriate marketing strategies in advance.

4. Application of data analysis

4.1. E-commerce

Big data analysis has helped e-commerce businesses sustain and grow their income while also improving their internal operations. Large data Analytics (BDA) is a method of large data analysis that gives decision-makers a view of the future. BDA has a variety of effects on e-commerce activities, including recognizing consumer happiness and analyzing user behavior and interest. For example, the user's emotions could be inferred from mouse cursor movements [12]. Through cluster analysis, an e-commerce company found that their customers can be divided into three segments: value, brand and discount. Value customers pay attention to cost performance, brand customers pursue high-quality brand products, and discount customers pay attention to price concessions. According to the characteristics of these segments, the company can develop different marketing strategies, such as giving discount to value customers, giving brand customers brand cooperation opportunities, etc.

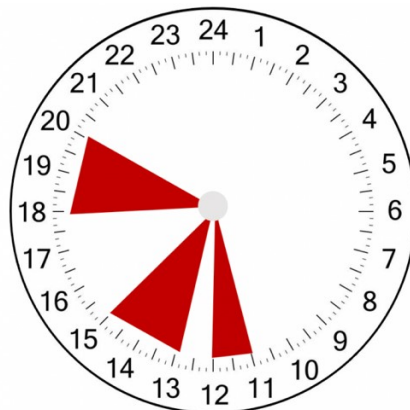


Figure 1: Prime time for online shopping. Customers typically shop between the hours of (a) 11:00 AM and 12:00 PM, (b) 1:00 PM and 3:00 PM, or (c) 6:00 PM and 8:00 PM [12].

The early buying time could be used to describe a logistical plan that delivers morning orders the same day. The busiest times for shopping are shown in Figure 1. Additionally, women shop

differently than males. Men prefer to finish their shopping in the morning, although women typically shop anytime of the day, particularly in the afternoons. E-commerce merchants can timing their promotions based on analysis of consumer purchases. Additionally, e-vendors can profit from big occasions like the new year's celebration and Black Friday to launch deals and other activities that draw in clients. Additionally, by offering individualized services and goods, businesses can draw customers by understanding customers' requirements and behaviors [12].

4.2. Tourism

Due to rising disposable incomes and an expanding global population, one of the major industries in the world is tourism, which is continually expanding [13]. For example, Greece has a number of important industries that support the prosperity of the economy. 16.5% of Greece's GDP and 18.4% of all jobs are produced directly by the travel and tourist sectors, plus related indirect and induced contributions. Additionally, revenue from tourism-related activities helps finance a portion of the balance of payments' current account deficit. Between 2000 and 2011, the tourism account had an average surplus of 4% of GDP, while the entire current account deficit was 9.5% of GDP [14].

An online travel platform found through correlation rule analysis that customers often book car rental services when booking hotels. Based on this finding, the platform can proactively recommend relevant car rental services when customers book a hotel, thereby increasing customer conversion rate and order value.

The above cases illustrate the application of data analytic in the field of customer segmentation and personalization. To better understand client needs and offer customized products and services that will increase customer happiness and loyalty, businesses can employ data analytics.

5. Conclusion

All in all, big data provides a valuable opportunity for businesses to better understand customer needs and personalize marketing. Through customer segmentation, precision marketing and predictive decision-making, companies can improve customer satisfaction and sales performance. In today's marketing process, enterprises need to accurately divide and analyze the market, propose personalized marketing strategies, and build brand image. However, when using big data for customer analysis and personalized marketing, businesses must also consider the challenges of data quality and privacy protection. Only by making full use of the advantages of big data and solving the corresponding challenges can enterprises achieve the goal of efficiently using big data for customer segmentation and personalized marketing.

At the same time, personalized marketing strategy has become an important means for enterprises to stand out. Through detailed analysis and understanding of customers, companies can offer tailored goods and services that boost client happiness and loyalty. In the highly competitive market environment, a personalized marketing strategy can help enterprises achieve sustainable growth and establish a good brand image. Therefore, enterprises should pay attention to the implementation of personalized marketing strategy, continuous improvement and innovation in order to satisfy the unique requirements of consumers. Only through continuous exploration and innovation, optimization and improvement of strategies can we gain more market share and reputation in the fierce market competition and become a leading brand in the industry.

Acknowledgments

Mengzhen Xiao, Yongchao Xu, and Zekai Gao contributed equally to this work and should be considered co-first authors.

References

- [1] Mahdavinejad, M. S., Rezvan, M., Barekatain, M., Adibi, P., Barnaghi, P., & Sheth, A. P. (2018). *Machine learning for internet of things data analysis: A survey*. *Digital Communications and Networks*, 4(3), 161–175. <https://doi.org/10.1016/j.dcan.2017.10.002>
- [2] Johnston, M. P. (2014). *Secondary Data Analysis: A Method of which the Time Has Come*. *Qualitative and Quantitative Methods in Libraries*.
- [3] Wilk, M. B., & Gnanadesikan, R. (1968). *Probability Plotting Methods for the Analysis of Data*. *Biometrika*, 55(1), 1. <https://doi.org/10.2307/2334448>
- [4] Cooil, B., Aksoy, L., & Keiningham, T. L. (2008). *Approaches to Customer Segmentation*. *Journal of Relationship Marketing*, 6(3–4), 9–39. https://doi.org/10.1300/J366v06n03_02
- [5] Anshari, M., Almunawar, M. N., Lim, S. A., & Al-Mudimigh, A. (2019). *Customer relationship management and big data enabled: Personalization & customization of services*. *Applied Computing and Informatics*, 15(2), 94–101. <https://doi.org/10.1016/j.aci.2018.05.004>
- [6] Kim, S.-Y., Jung, T.-S., Suh, E.-H., & Hwang, H.-S. (2006). *Customer segmentation and strategy development based on customer lifetime value: A case study*. *Expert Systems with Applications*, 31(1), 101–107. <https://doi.org/10.1016/j.eswa.2005.09.004>
- [7] Fan, H., & Poole, M. S. (2006). *What Is Personalization? Perspectives on the Design and Implementation of Personalization in Information Systems*. *Journal of Organizational Computing and Electronic Commerce*, 16(3–4), 179–202. <https://doi.org/10.1080/10919392.2006.9681199>
- [8] Sundar, S. S., & Marathe, S. S. (2010). *Personalization versus Customization: The Importance of Agency, Privacy, and Power Usage*. *Human Communication Research*, 36(3), 298–322. <https://doi.org/10.1111/j.1468-2958.2010.01377.x>
- [9] Essalmi, F., Ayed, L. J. B., Jemni, M., Kinshuk, & Graf, S. (2010). *A fully personalization strategy of E-learning scenarios*. *Computers in Human Behavior*, 26(4), 581–591. <https://doi.org/10.1016/j.chb.2009.12.010>
- [10] Kruschke, J. K. (2010). *What to believe: Bayesian methods for data analysis*. *Trends in Cognitive Sciences*, 14(7), 293–300. <https://doi.org/10.1016/j.tics.2010.05.001>
- [11] Kruschke, J. K., Aguinis, H., & Joo, H. (2012). *The Time Has Come: Bayesian Methods for Data Analysis in the Organizational Sciences*. *Organizational Research Methods*, 15(4), 722–752. <https://doi.org/10.1177/1094428112457829>
- [12] Alrumiah, S. S., & Hadwan, M. (2021). *Implementing Big Data Analytics in E-Commerce: Vendor and Customer View*. *IEEE Access*, 9, 37281–37286. <https://doi.org/10.1109/ACCESS.2021.3063615>
- [13] Poprawe, M. (2015). *A panel data analysis of the effect of corruption on tourism*. *Applied Economics*, 47(23), 2399–2412. <https://doi.org/10.1080/00036846.2015.1005874>
- [14] Chasapopoulos, P., Butter, F. A. G. D., & Mihaylov, E. (2014). *Demand for tourism in Greece: A panel data analysis using the gravity model*. *International Journal of Tourism Policy*, 5(3), 173. <https://doi.org/10.1504/IJTP.2014.063105>