Big Data Analysis for Consumer Behavior: Evidence from User Segmentation, Recommendation and Opinion Monitoring.

Quanfu Chen[1], Qianru Lei[2, *] and Ailun Zheng[3]

[1]Zhejiang Ninghai High school, Ningbo, China

[2]Shenzhen International Foundation College, Shenzhen, China

[3]Ulink College of Shanghai, Shanghai, China

*Corresponding author: s.chihiro.lei@sifc.net.cn

Abstract. As a matter of fact, digitisation in the 21st century has brought with it an infinitely growing volume of data in recent years. On this basis, the cost of storing data has become less expensive than cleaning it, marking the dawn of the Big Data era. On this occasion, for consumer behavior, big data analytics (BDA) can greatly improve the efficiency of sub-consumer behavior analysis as well as the volume of data, making the research results more accurate. With this in mind, this paper will analyze the applications and implementations of bigdata analysis in the field of consumer behavior. To be specific, this paper is divided into three main parts: first, a brief overview of the background of consumer behavior and BDA; second, a detailed description of the three BDA applications in consumer applied science, and finally, an outlook on the future of BDA in this field. Overall, these results shed light on guiding further exploration of consumer behavior analysis.

Keywords: Big data analytics; consumer behavior; user segmentation; opinion monitoring.

1. Introduction

It is clear that the consumer behavior has already contributed to many fields, such as finance, economics, big data science. Before the deeply study, we ought to know the history of the development of this subject [1]. Seen from Table 1, consumer behavior is a recent birth of the discipline which is less than a hundred years old with the conforming, exploratory, and of long-term and far-reaching significance for the society. Throughout history, the role of consumer behavior is mainly focused on the consumer analytics related to the big data. More deeply, the key is between the marketing managers and decision makers, in order to achieve sustainable business growth [2]. In addition, the consumer behavior are widely applicated in the development of financial services model [3]. Basically, these are the contribution of the consumer behavior theory and the meaning of research.

Table 1. Timeline of the development of the subject of the consumer behavior.

Time	Events
1940-1950	The appearance of the consumer motivation research
1963	The appearance of a concept: “lifestyle”
1968	The publication of “Consumer Behavior” by Blackwell and Miniard
1969	The proposal of theory of consumer behavior by Howard and Sheth
1974	The start publication of Journal of Consumer Research (JCR)
1982	Self Concept is introduced into consumer behavior
1985-1991	The appearance and perfection of the theory of planned behavior (TPB)
1992	The start publication of Journal of Consumer Psychology (JCP)
1995	The publication of the book “Being Digital” by Nicholas Negroponte
2001	The proposal of the concept of the brand community
2002	The start publication of Journal of Consumer Behavior (JCB)
2004	The appearance of the social media: Facebook
2005	The proposal of the concept of consumer culture theory
2010	The concept and the research of the digital social media and the digital consumer behavior appear

We are living in the era of the information revolution. The term big data spans a wide range of industries, from meteorology to biological research, from finance to healthcare, big data is being

applied in many scenarios [4]. Because of the challenges and opportunities produced by the information revolution, big data analytics (BDA) has grown as a fresh field of innovation across the entire spectrum of the electronic commerce scene [5]. Companies can obtain a massive amount of data from a variety of sources, including online conversations, user transactions, purchasing patterns and consumer profiles. Because the sheer amount of data grows drastically, "Big Data Analysis," a new arising technology, is being researched and used for analyzing consumer behaviour.

Fig. 1 depicts a simple step-by-step methodology for BDA that uses collation, pre-analysis, analysis, modeling, and prediction to infer customer behavioural trends from raw data. In real life the work is more complicated. Big data research enables firms to identify specific consumer preferences, purchasing trends, and behaviours. The resulting consumer behavioural information can be used to improve the shopping experience in e-commerce projects and raise income for suppliers. Furthermore, understanding client demands and behaviour enables businesses to attract customers by providing customised offerings [6]. As it stands, employing big data to analyse customer buying patterns has become increasingly crucial for organisations seeking to keep themselves competitive in the contemporary marketplace [7]. For example, Amazon applies BDA to redevelop its distribution strategy by forecasting customer purchasing behaviour and transporting goods to the appropriate warehouse in advance. Consequently, Amazon generates more value with this technique [8]

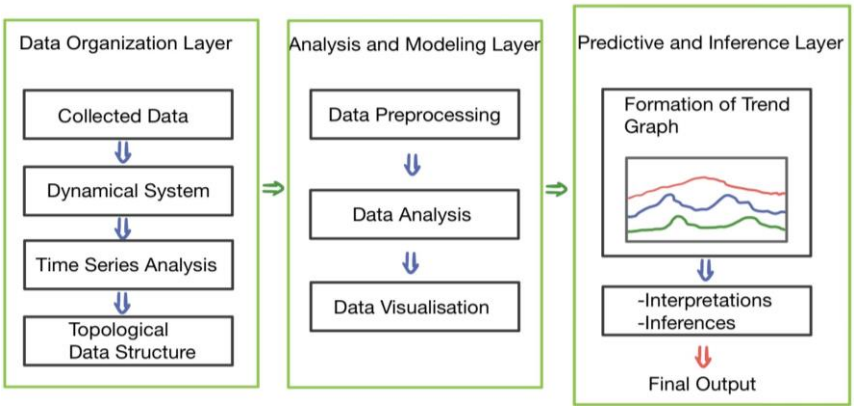


Fig. 1 Big Data application framework for consumer behaviour

BDA can help researchers reveal the purchasing decision-making process of consumers. For example, by analyzing consumers' browsing history and purchase history on e-commerce platforms, we can infer consumers' preferences for different categories of goods and optimize product recommendation and pricing strategies based on these results. Second, BDA can help researchers understand consumers' shopping behavior patterns. By analyzing consumers' behavior data in the shopping process, we can reveal consumers' purchasing habits, shopping channel preferences and transaction preferences. These data can provide enterprises with valuable market insights, help enterprises optimize channel layout and promotion strategies, and improve sales efficiency. Third, BDA can help researchers identify the psychological characteristics of consumers. By analyzing the speech and behavior of consumers during the shopping process, we can study the emotional feedback and psychological state of consumers. This is very important for businesses, because understanding consumers' emotional feedback can help companies better understand market needs, improve product design, and increase user satisfaction. Finally, BDA can help researchers predict consumers' future purchasing behavior. By building predictive models, we can use consumers' historical data and behavior patterns to predict their future buying intentions and purchasing power. This is very valuable for businesses because they can market and allocate resources based on the predicted results to obtain higher sales revenue.

Understanding consumer behaviour is now a top strategic priority for businesses, and those striving to remain competitive must leverage BDA to analyse consumer preferences and improve engagement. The goal of this paper is to demonstrate the immense potential of BDA in unravelling consumer behaviour, discussing and analysing its ability to provide accurate user segmentation, make purchase predictions and brand sentiment analysis. Since BDA research is still in an emergent field, researchers

consistently investigating new techniques, methodologies and continuously exploring the potential of the area of BDA. Therefore, this field is still evolving and a search period spanning from 2015 to 2023 was deemed representative for this paper.

2. Basic Descriptions

BDA is the process of analyzing and comprehending massive amount of data, such as emails, comments on social networks and so on, which attempts to optimize decision-making. Trends and connections can be detected in enormous amounts of raw data by the steps as shown in Fig. 1. BDA in the field of consumer behavior can provide firms and researchers with a comprehensive understanding to assist in understanding consumer preferences, requirements, and behavior. Several brands that market on social media have to employ various platforms to post adverts and to make connections with their customers in order to brand and enhance site visitor density and income [9]. Therefore, they can utilize BDA to determine what type of material to post that will be more attractive (e.g., LinkedIn, Facebook, Instagram, Twitter, YouTube). There are numerous technologies that can be used in BDA, including API, stream analytics, web analytics, etc., that support the tracking, processing, analysis, and prediction of raw data [10]. One of the technologies is mobile analytics, which aims to retrieve useful insights and trends from unprocessed mobile data [11]. It provides insights on consumers' mobile behavior patterns by analyzing clickstream data logs and sensor data (such as audio sensor data and location data) generated by phones and tablets. Analyzing website visitor behavior is done using web analytics. The raw data gathered could consist of view counts, routes taken, and the rate of conversion. By personalizing services for frequent visitors, tracking the correlation between purchases of particular goods and locations, and reestablishing website pages that are most probable stimulating purchases, the information obtained can help businesses better understand the purchasing process of consumers and increase turnover. SMA helps businesses to better comprehend consumer activity and preferences and predict consumer behavior by analyzing social media data like user postings, conversations in the comments, follow lists, etc.

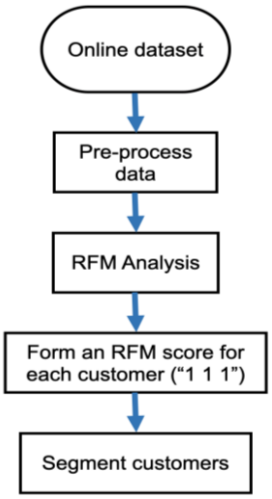


Fig. 2 The RFM model's framework.

3. User Segmentation and Customised Marketing

In traditional retail industry merchants use descriptive method, which is lacking in market understanding [12], i.e., record how many of which products were sold at the end of the day. Businesses then began to consciously try to understand consumers, including their needs, preferences, and habits, in order to provide more suitable products to meet their demand. Prior to the 1990s, businesses implemented the results of consumer questionnaires, and geographic, demographic and audience information from different forms of media to proceed with market segmentation. At this moment, businesses are gradually moving away from mass marketing and towards customized

promotion, or "one-to-one" advertising [13]. With the advent of the internet and the soaring amount of data, for the sake of providing personalized marketing and targeted advertising, companies began to use BDA in models such as RFM and CRM to segment customers and markets (seen from Fig. 2 and Table 2). While it is significant for businesses to provide tailored advertisement to potential customers, it is also important to retain existing customers. Businesses want to be able to categories customers according to their purchasing history (Recency, Frequency and Monetary) and differentiate their marketing strategies in order to enhance their customers' brand stickiness. They may quickly identify important clients and adjust their marketing tactics according to the result calculated by the RFM model. It is used in a variety of domains, particularly for Direct Marketing (including marketing via phone calls, social media, and e-mail). The RFM model is broadly utilized in comparison to other algorithms since it employs fewer parameters for segmentation [14]. Depending on their values of Recency, Frequency, and Monetary, after the RFM algorithm, each purchaser is labelled 555, 554...111. Score 555 means that this potential customer has a great possibility of providing greater benefits to the company, and the situation is the opposite with 111 [15].

Table 2. RFM score description.

Score	Characteristics
5	Potential
4	Promising
3	Can’t lose them
2	At risk
1	Lost

Businesses need to preserve an ongoing relationship with their consumers because it is generally more expensive to acquire new customers than to keep existing ones [16]. The firm platform and the CRM system can be connected, enabling real-time updating of customer data. Data mining is one use of BDA that is frequently employed in CRM nowadays. Descriptive analysis, diagnostic analysis, predictive analysis, and perspective analysis are the four processes that make up the CRM process based on big data [17]. These analytics are employed to recognize and comprehend present consumer behavior, forecast future developments, and make suitable marketing decisions for businesses. Consumer behavioral data can be analyzed through BDA to understand different customer segments. Consumer segmentation enables businesses to tailor their marketing strategies to specialized groups of consumers, placing advertisements and promotions that meet consumers’ unique needs and preferences. This customized strategy has the characteristics of being efficient and profitable, which is more beneficial than a generic strategy.

4. Purchase Prediction at E-commerce

With the advent of the era of big data, more and more enterprises are beginning to understand the importance of data, especially in the field of e-commerce. Online shopping has become one of the main ways of modern consumption. Under this background, the intelligent data processing technologies on ecommerce users’ purchase prediction are showing their importance, such as Collaborative Filtering [18]. The research on consumer purchase prediction and recommendation system of e-commerce platform is to mine the rules and characteristics of consumers' shopping behavior through BDA and machine learning methods, predict consumers' future shopping behavior, and provide personalized product recommendation services for different users. The goal of the recommendation system is to improve consumers' shopping satisfaction and purchase conversion rate, while also improving sales and user loyalty of the e-commerce platform. Here are some ways that companies can use recommendation systems to improve the accuracy of predictions about consumer buying behavior:

-Data collection and anaylisis

-Personalized recommendation algorithm

-Feature engineering
-Real-time recommendation and feedback
-A/B testing and evaluation

Consumer demand forecasting is an important application of e-commerce forecasting model. Through data collection and analysis, the e-commerce forecasting model based on big data can accurately predict the demand trend of consumers, which plays an important role in grasping market changes and formulating enterprise strategies. Commonly used prediction models are mainly divided into the following categories：

-Rule model: Using statistical regression method to calculate the basic sales, and then superimpose various coefficients, such as festivals, marketing, seasonality, etc.
-Classical time series models: ARIMA, STL, etc.
-Decision tree model: GBDT
-Deep neural networks: DeepAR, WaveNet, Transformer, etc.

5. The sentiment analysis and the public opinion monitoring.

Fundamentally, this contribution is denoted on the organizations in the field of finance or business to utilize slant investigation for the dynamic of their brands [19]. In order to achieve the goal, the organizations may use AI as a tool to disposal and classify the data. AI can quantify these information from unpredictable messages to the measurable data. Thus, based on the precious limited data, AI is capable to predict the tendency [20] by assembling information into Classification, Regression, Clustering, Anomaly location, Recommendation, and Dimensionality decrease. [21]. There is a fundamental and wide-used distributed steps as a arithmetic as illustrated in Fig. 3.

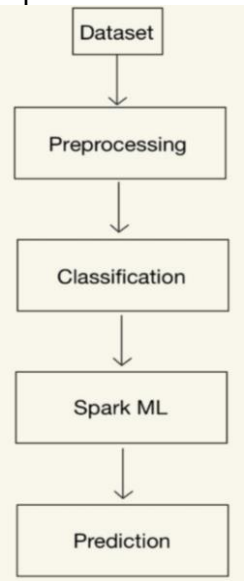


Fig 3. The process of analysis data

The dataset means the collection of data by administers AI calculations. It is clear that by automatic machine calculation, the results may be more accurate and faster than manual operations. For the processing, firstly, the feature extraction is usually the term recurrence reverse archive recurrence (TF - IDF), which is a capacity vectorization instrument. Then, for the feature transformers, tokenization is a useful way, is a converter that naturally changes the given information string to lowercase, and by utilizing whitespaces, sprinkles the string into characters, e.g., sprinkles the info string into a grouping of characters [22]. For the classification, it is usually the method of Gaussian Native Bayes Classification, whose aim is to place a kernel function on each training data point for regression each point. Thus, this framework extremely speeds up the collection and classification of data [23]. Spark ML, which is a popular open-source platform for large-scale data processing that is well-suited for the study of consumer behavior theory under the big data. We present MLlib, Spark’s

open-source distributed machine learning library. The reason we use MLlib is its efficient functions to lots of fields such as statistics, algebra and linear function [24]. One example of ML Algorithms is the Classification Algorithms (GNB and RF). The code will be written in the Python. Next, input the relative information index of different public opinions. Hence, the results will be presented by the whole procedure. 'MLlib exhibits excellent performance and scalability' [24], and that is the reason we use it. After these steps, it is necessary to test its accuracy by Python 3.0. In order to make our results more convince, we get another results of Random Forest (RF) from Python as well. The results show in Table. 3. As the table shows, by the GNB, our results turn more accurate and precise, in comparison with the RF. For the prediction, based on the relative accurate classifier, it is possible to foreshadow the next state consumer (sentiment in this case). If it is more rigorous, sometimes these data will be recalculated by a system of Markoff Chain. In conclusion, based on the whole procedure: from dataset to prediction, the MLlib and the GNB, which are both on the basis of big data, made a great contribution. Even later on standpoint, this procedure can be utilized in more categories of information indexes to foresee some specific data.

Table 3. The table of accuracy classifier.

Accuracy Classifier	GNB	RF
1	0.7698	0.7186
2	0.7701	0.7211
3	0.7723	0.6993
4	0.7716	0.7057
5	0.7729	0.7327
6	0.7741	0.6881

6. Limitations and Prospects

The limitation of big data's responsibility. The big data collected by AI rather than human is unemotional and unsentimental. Thus, the responsibility for the data collection is a big problem, yet it is difficult to hold algorithms accountable as they continuously evolve with technologies, systems, data and people, the ebb and flow of policy priorities, and the clashes between new and old institutional logics [25]. Whether it is the application 1, 2 or 3. Thus, this potential moral challenge may limit the widespread of the application of big data. The limitation of big data's practicability. Always, an effective data analysis should be under a huge database. However, in the reality, the data based on the current situation of the world's consumer are few, due to the capricious changing situation of consumer. In other words, lots of old data - not in keeping with the current situation- are useless. Thus, it is hard to expect based on the real-time cases. This model is assumed that every consumer is rational. Then, AI analyze data based on this assumption. However, in the reality, people are not always rational. For instance, the consumer's decision for choosing a goods may influence by the sunk cost they have done for this good. In the field of government affairs, big data can realize the sharing, integration and opening of government information resources, improve the level of government services and public decision-making capabilities, and build smart cities and digital governments. In the financial field, big data can realize the identification, prediction, prevention, and control of financial risks, improve the innovation and personalization of financial products, and build inclusive finance and digital finance. In the field of e-commerce, big data can realize the analysis, mining and prediction of consumer behavior, improve the effect of product recommendation and marketing, and build precision marketing and smart retail; In the medical field, big data can realize the optimal allocation of medical resources, improve the quality and efficiency of medical services, and build smart medical and digital health.

7. Conclusion

To sum up, the emergence of BDA has brought about a paradigm shift in how to understand consumer behavior and address their needs. This paper explores the various impacts of BDA on consumer behavior, such as helping companies to customize marketing, purchase prediction and analyze customers' sentiment towards brands. Furthermore, this article introduces different consumer behavioral models that involve BDA such as RFM CRM and prediction models, showing in detail how BDA can be used in consumer behavior and the great potential of this tool in this field.

Author Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] Lu T. 50 Years of Conusmer Behavior: Evolution and Overthrow, Foreign Economics & Management, 2017, 39(6): 24.

[2] Md Afnan H, Shahriar A and Venkata K Y. Customer Analytics Capabilities in the Big Data Spectrum: A Systematic Appr Systematic Approach to Achie o Achieve Sustainable Firm PE. Sustainable Firm Performance, 2020, 1-1: 9-10.

[3] Milner T, Rosenstreich D. A review of consumer decision-making models and development of a new model for financial services. GSB, 2013: 106-120.

[4] Yin S, Kaynak O. Big data for modern industry: challenges and trends. Proceedings of the IEEE, 2015, 103(2): 143-146.

[5] Akter S, Wamba S F. Big Data Analytics in e-commerce: A systematic review and Agenda for Future Research', Electronic Markets, 2016, 26(2): 173–194.

[6] Alrumiah S S, Hadwan M. Implementing big data analytics in e-commerce: Vendor and Customer View. IEEE Access, 2021, 9: 37281–37286.

[7] Yangzom T, Ahuja L. Uses of big data to understand consumers' buying behavior', 2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC), 2023: 101-102.

[8] Erevelles S, Fukawa N, Swayne L. Big Data consumer analytics and the transformation of marketing. Journal of business research, 2016, 69(2): 897-904.

[9] Chaudhary K, Alam M, Al-Rakhami M S, et al. Machine learning-based mathematical modelling for prediction of social media consumer behavior using big data analytics. Journal of Big Data, 2021, 8(1): 1-20.

[10] Lehrer C, Wieneke A, Vom Brocke J A N, et al. How big data analytics enables service innovation: materiality, affordance, and the individualization of service. Journal of Management Information Systems, 2018, 35(2): 424-460.

[11] Alsheikh M A, Niyato D, Lin S, et al. Mobile big data analytics using deep learning and apache spark. IEEE network, 2016, 30(3): 22-29.

[12] Yoseph F, Ahamed Hassain Malim N H, Heikkilä M, et al. The impact of big data market segmentation using data mining and clustering techniques. Journal of Intelligent & Fuzzy Systems, 2020, 38(5): 6159-6173.

[13] Payne A, Frow P. Customer relationship management: Strategy and implementation. The Marketing Book. Routledge, 2016: 439-466.

[14] Rezaeinia S M, Rahmani R. Recommender system based on customer segmentation (RSCS). Kybernetes, 2016, 45(6): 946-961.

[15] Christy A J, Umamakeswari A, Priyatharsini L, et al. RFM ranking–An effective approach to customer segmentation. Journal of King Saud University-Computer and Information Sciences, 2021, 33(10): 1251-1257.

[16] Law R, Fong D K C, Chan I C C, et al. Systematic review of hospitality CRM research. International Journal of Contemporary Hospitality Management, 2018, 30(3): 1686-1704.

[17] Perera W K R, Dilini K A, Kulawansa T. A review of big data analytics for customer relationship management. 2018 3rd International Conference on Information Technology Research (ICITR). IEEE, 2018: 1-6.

[18] Zheng Y, Liu C. Neural Autoregressive Collaborative Filtering for Implicit Feedback. Proceedings of the 1st Workshop on Deep Learning for Recommender Systems. Boston, MA, USA, ACM., 2016: 2-6.

[19] Al-Barznji K, Atanassov A. A framework for cloud based hybrid recommender system for big data mining. a journal of "Science, Engineering & Education, 2017, 2(1): 58-65.

[20] Benjamin B, Jenny K. Data Analytics with Hadoop. O'Reilly Media, Inc., 2016.

[21] Guller M. Big data analytics with spark. Springer, 2015.

[22] Rekha K B, Naveen Chandra G. A framework for sentiment analysis in customer product reviews using machine learning. 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE) 2020: 268-270.

[23] Williams C K I, Barber D. Bayesian classification with Gaussian processes. IEEE Transactions on pattern analysis and machine intelligence, 1998, 20(12): 1342-1351.

[24] Meng X, Bradley J, Yavuz B, et al. Mllib: Machine learning in apache spark. The journal of machine learning research, 2016, 17(1): 1235-1241.

[25] Janssen M, Kuk G. The challenges and limits of big data algorithms in technocratic governance. Government Information Quarterly, 2016, 33(3): 371-377.