

# ***Implementation of Bigdata Analysis in Consumer Behavior***

**Bonian Han<sup>1,a,\*</sup>, Ziming Xiong<sup>2,b</sup>, Xiaohe Xu<sup>3,c</sup>, Yuchi Zhang<sup>4,d</sup>**

<sup>1</sup>*School of Electronic Information, Hangzhou Danzi University, Hangzhou, China*

<sup>2</sup>*SWUFE-UD Institute of Data Science, Southwestern University of Finance and Economics, Chengdu, China*

<sup>3</sup>*Tianyi high school, Wuxi, China*

<sup>4</sup>*St. George College, The University of Toronto, Toronto, Canada*

*a. 22040120@hdu.edu.cn, b. 676000098@qq.com, c. todayisfriday2021@163.com,*

*d. 1843342040@qq.com*

*\*corresponding author*

**Abstract:** Nowadays, as the way of bigdata analysis become more and more diverse and specific, some people have already turned their eyes to the implementation of these analysis in consumer behavior. Since the markets are more competitive, it would save time and money that the companies produce what the consumers like. In addition, many markets exist for a long time and companies collect a plenty of consumers' data, when they use bigdata analysis on the data collected, they can clear make accurate prediction about future products. In this study, we split the implementation of bigdata analysis in consumer behavior into several parts, including the history of the research, the specific analyzing methods, the realistic applications and limitations. In each part, we combine the facts and the understanding to write the analysis by the reference of some authoritative documentations. As a matter of fact, there are two significances of research, first is to have a comprehensive understanding of the current situation about topic from by-parts investigation; second is to have a future imagination based on both the advantages and disadvantages have now.

**Keywords:** Consumer behavior, big data, data analysis

## **1. Introduction**

In the era of big data, big-data analysis has penetrated all levels of society. Especially among the participants in the network economy, bigdata is playing an increasingly important role. Investigation on Big-data analysis in consumer behavior mainly started in 83 years ago, by Foxall [1]. The emergence of consumer behavior research as a social science can trace its direct roots back to the 1940s, when it grew out of the marketing discipline to also incorporate the study of psychology, anthropology and economics. At first, big-data analysis was formed because of the increasingly intense competition based on the Internet be extensively used. Big-data analysis was used for increasing the sales and produce with a certain direction. The big-data consumer research was initially aimed at Big improving customer relationships and offer products or services that best suit their needs.

After the development from 1940, there are 4 effects that big-data analysis of consumer behavior brings to the society. There are some perspectives concluded by several internet enterprises. It is conducive to enhancing the competitiveness of enterprises. Besides, it is conducive to correctly

guiding consumer demand. Thirdly, it is conducive to the formulation of macroeconomic policies and laws. In history, big-data analysis also has developed several steps of main calculation and prediction. Those effects promote the development of whole society.

Nowadays, big-data analysis has been extensively used. Study shows big-data analysis is not only used for sales recently but also for other areas. Field like education, healthcare and government sectors all accept the high technology data dealing and calculating progress. For example, in education, bigdata can customize and dynamic learning programs [1]. By analyzing the detailed data and experience in the past, with the help of statistic, bigdata contributes to diversification. However, though the bid-data is widely accepted in this era, it still faces a lot of challenges even in its use or future development. Many people are worrying about the negative outcomes of big-data analysis, about the privacy of life or whether they have been under the monitor. There is some information of a person which when combined with external large data may lead to some facts of a person which may be secretive, and he might not want the owner to know this information about that person [2]. Moreover, another challenge is that the amount of data processed daily around the world is constantly increasing, which makes big data a relative term [3].

Nowadays, some new technologies such as cloud computing and big data always hope that whenever a failure occurs, the damage caused should be within an acceptable threshold, that is, the entire task should not be started from scratch. It also requires that system failures be handled in an efficient manner. This brings up the big question of which storage device to use. In this study, we put effort to the study of how big-data analysis use statistic models to calculate the accurate data and make our prediction. Secondly, in difficult and complicated situation, how does the model work different from the normal situation.

## 2. Basic Descriptions of Bigdata Analysis

In the realm of contemporary data-driven enterprises, the analysis of vast and intricate datasets, commonly referred to as big data, has emerged as a pivotal driver for extracting invaluable insights and knowledge. The multifaceted nature of big data, distinguished by its substantial capacity, rapidity, and diversity, necessitates the application of intricate methods and models to unearth meaningful patterns and trends. In the initial phase of big data analysis, data preprocessing and cleaning play pivotal roles in ensuring the quality and reliability of subsequent analysis outcomes. This phase encompasses techniques such as data normalization, transformation, and outlier detection, which are aimed at eliminating noise, inconsistencies, and irrelevant information. The objective is to enhance the accuracy of subsequent analytical models by cultivating a refined dataset.

Descriptive analysis serves as the cornerstone of the analytical journey, encompassing the summation and visualization of data to discern potential patterns and features. employing statistical tools and visualization techniques aids in comprehensively portraying data, fostering a comprehensive grasp of its dynamics, and facilitating the identification of preliminary insights. Predictive analysis harnesses the potency of historical data to endeavor to predict forthcoming events or trends. By employing an array of statistical methods and machine learning algorithms, prediction models such as regression, decision trees, and neural networks are constructed to anticipate impending outcomes. The intricate interplay between past data and advanced algorithms empowers organizations to make informed predictions, thereby amplifying strategic decision-making. Prescriptive analysis represents the zenith of data-driven decision-making, offering optimal actions to steer outcomes towards predefined objectives. The fusion of historical data, predictive insights, and business rules culminates in recommendations that maximize anticipated results. This paradigm empowers organizations not only to prognosticate future events but also to proactively influence them in a manner conducive to their aspirations. Delving into the nascent prospects of text data involves employing text and sentiment analysis, leveraging natural language processing (NLP) technology to

extract meaningful insights from textual sources. By discerning emotions, identifying pivotal themes, and uncovering textual patterns, this analytical facet enriches analysis with linguistic dimensions, enabling organizations to comprehend the intrinsic unstructured narrative of contemporary data streams (seen from Fig. 1).

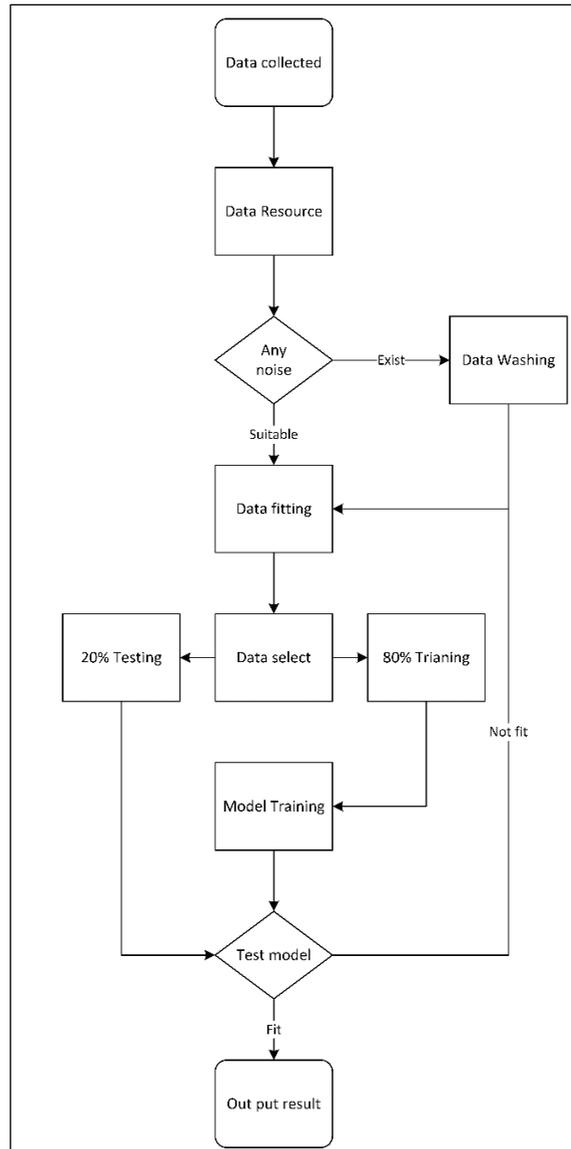


Figure 1: The processes of training data.

Clustering technology plays a pivotal role in unearthing latent structures in datasets and facilitating the identification of groups of similar data points. The application of methods such as K-means clustering, and hierarchical clustering facilitates data segmentation, focused analysis, and the revelation of latent intricate relationships. Through mining association rules and scrutinizing graphs, the exploration of associations and relationships within the dataset is fostered. These methodologies unveil intricate patterns and connections between variables or entities, laying the groundwork for the utilization of market basket analysis, social network analysis, and recommendation systems. Machine learning forms the bedrock of big data analysis, encompassing a myriad of algorithms such as decision trees, support vector machines, random forests, and deep learning. Through iterative learning from data patterns, machine learning models can classify, predict, and distinguish complex

relationships, thereby enriching analysis outcomes. The enormity of big data necessitates the adoption of distributed computing frameworks such as Hadoop and Spark to accomplish parallel processing across distributed nodes. This approach adeptly addresses computational requisites tied to sizable datasets. Additionally, real-time analysis conducted through platforms like Apache Kafka and Flink empowers organizations to swiftly glean insights from data streams, thus fostering timely decision-making. Effectively conveying analysis results to stakeholders calls for advanced visualization techniques. By means of dashboards, heat maps, and interactive visualization, the intricacy of analysis outcomes is distilled into a comprehensible format, empowering stakeholders to promptly absorb insights and make well-informed decisions.

### 3. Applications in Social Media

In 2021 Kiran et al. construct a model to analysis the behavior of every social media platform based on big data and machine learning [5]. They referenced a study involving the analysis of non-structured correlated big data, and in this study used at least 95% of total data [6]. They were motivated by an article on a machine learning-based approach to enhancing social media marketing. Thus, in this era and environment as a big background, there a huge amount of data could be provide as source, to build, training and forecasting the behavior that what consumers will interesting on something or buy something.

The model they build was based on a liner regression model and Decision Tree Regressor, Random Forest Regressor, and so on. As most machine learning models are built, the 80% of data they collected was taken as training data, and the 20% left as the testing data. Before data input the model, the first step is clean the data. Removing the dirtiness will make the data acquire higher quality to make model more accuracy, like remove the outliers, noises and errors. In this step some common methods for removing data noise and errors include like using statistical methods such as stand deviation, box plots and some other method to identify the outline data, then it can be replaced by the mean, median or using interpolation method to fill this blank. These are the simplest, most understandable, and commonly used methods. After the cleaning step, they using the liner model to perform data fitting, and find features from data also represent the consumer behavior, then get the output function that could be used to predict the user preferences under the influence of varying factors.

And, in former research [7], they assessed consumer allocation to social media, which to some extent amalgamates consumers' responses to marketing communications. In this study, the authors' argument posits that the certainty motivation underpinning social media usage contributes to being a precursor to the general disposition towards social networking sites, thus consecutively influencing the attitude towards sellers on social networking platforms. Obviously, the author cites this article and to help build and improve their model to calculate what customers like and dislike, the conclusions are extremely close.

For this social media case, the most suitable model they concluded was Decision Tree Regressor model, it could help operation managers forecast their customers next behavior more precisely, like based on the data that what consumer prefer, who do they follow, and what they download. These will help mangers make their production more in line with consumer preferences and improve competitiveness with other opposites.

### 4. Applications in Low-carbon Tourism

In addition, not only the simple model like some liner model, but also in some complex situations, the forecast will be more difficult. Another case created by Ma et al.hey want to investigate the utilization of big data to promote low-carbon smart tourism, with a specific focus on consumer

behavior and corporate altruistic preferences within the context of the low-carbon tourism online-to-offline (O2O) supply chain. Compare with the former case of social media, this will be more realistic and more significant. In a rapid development period, consumer's consciousness of protect environment is growing rapidly as well. Now, low-carbon consumption has become a consensus, an increasing number of companies are incorporating the low-carbon concept, integrating the notion into their operational frameworks [9]. The author builds a supply chain that named the Low-carbon smart tourism supply chain. It includes three participants: tourist scene (TS), online travel agent (OTA) and Tourists. So, the article aims to address the issue of promoting the sustainable and intelligent development of the low-carbon tourism supply chain through the utilization of big data marketing and corporate decision-making. The author based on the big data-empowered low-carbon tourism supply chain, and for making different strategies, the differential game models were established for solving the low-carbon goodwill time trajectory, enterprise profit, and few other problems. A sketch is shown in Fig. 2. After the analysis process, for this problem, the article underscores the significance of consumer reference effects and corporate altruism in promoting low-carbon goodwill and environmental benefits. The result showed that the decision mode as a significant factor that influencing the corporate performance. Furthermore, author use the numerical examples to test the analytical outcomes.

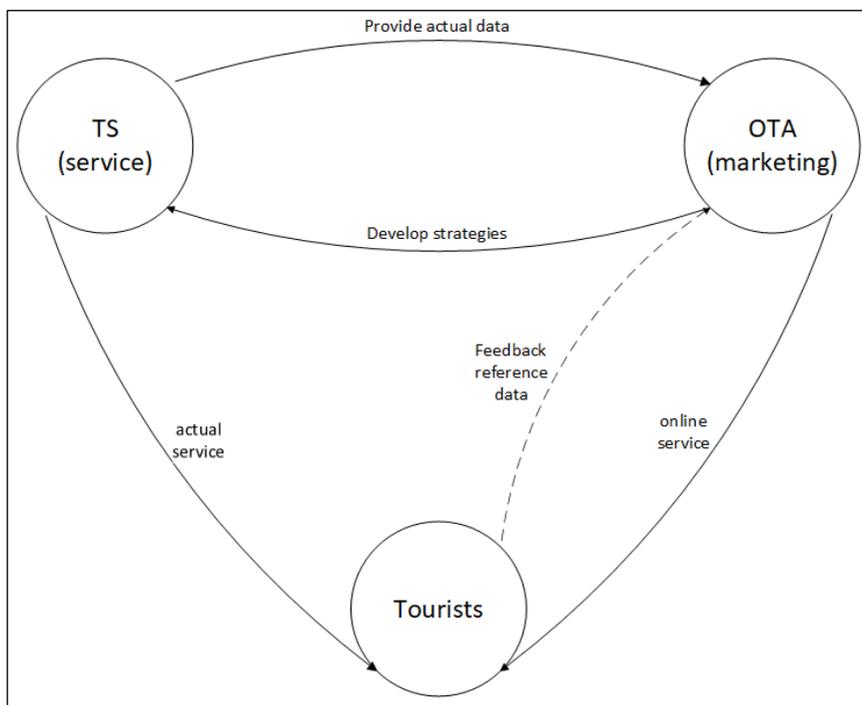


Figure 2: The relationship between TS, OTA & Tourists

## 5. Applications in B2B

In another perspective, Holland gave a new method to analysis the consumer behavior, and the mean application is to build a network that connect the competitors, search intermediaries and consumers [10]. And they construct a From-to Diagramming to illustrate the flows that loss and gain of an individual website to other websites. It could see the relation of the flows change among different locations, and use the Network Diagramming with different color to show the strong correlation and weak correlation, then the visualization of the online market will help the company to make the suitable strategy and auxiliary analysis intuitively. If the manager wants to dig something deeper,

using algorithm will be a necessary method to calculate and analysis, then managers can rely on these results to make some relevant predictions. In this case, the author builds several index to standardization and separate the different location, and gave them different evaluate score. The data by computed will show the performance of the relationships between the individual airlines and the other airlines as a group of websites. In another algorithm they use a similar way to detailed online performance, its main purpose is to show the composition of this average performance and illustrate a large variance of scores with online travel agents. In other words, it could identify a potentially fruitful area for further market research, and it possibly depend to the segmentation of customers using different OTAs.

This typical B2B model is also very common in all kinds of business strategy formulation, but combine it with big data and algorithms, the result will be more significant and direct. Even some managers have little known with big data, but the analyzed data will give the significant effect and feature of consumers behavior obviously. Then with this information, enterprises could make more products and more competitive strategies to adapt to customer requirement and finally get rewarded in the market.

## 6. Limitations and Prospects

For the limitation of the implementation of bigdata analysis on consumer behaviors, according to Bormida [11], Huang et al. [12], Liu et al. [13], the data privacy and security problem is the most important limitation currently have during using the bigdata. At most of the time, when consumers trading or using the tools on the internet, it will form many personal data about the consumers. Sometimes, when the companies or other institutions use bigdata to analyze the consumer behavior, they will extract this information, and based on the high deduction ability, it is easy for the companies to determine the personal thoughts and personal tendencies of the behavior, some of the tendencies or thoughts are even what the consumers tend to avoid. The reason why this situation happens many times is due to the idea that “traditional methods and notions of privacy protections might be inadequate in some instances” by Bormida.

Another limitation is the quality of the data used for analysis, sometimes, some companies prefer to do the analysis based on the original data provided without any check or filter. In fact, as the data on the internet become more and more, without further checking or filtering, it is hard to make sure the accuracy and comprehensiveness of the analysis. There always exist problems like data falsify or manipulation, these data including products description, image showing or even the credit index [11], those problems clearly effect the feedback of consumers and thus effect the analysis. Since companies, consumers and data provided form a cycle, when the data to both sides are falsified, the whole cycle will be influenced, even broken.

For the future outlook, it based on two parts, one is the fixation of the existing limitations and the other is development of some good features. There are several steps can be done for the fixation, first it is important for the government to set up new policies to further ensure the future data security, including set detailed rule about what kind of data is secure and can be used to do the analysis and when the companies or other institutions want to use the amount of data that beyond some level, the data collection website need to send a notification to the specific consumer in order to receive the further permission about the use of data. Second is raise the level of punishment, keep punishing the companies or individuals who utilize the data privately and collect the data that consumers prefer to hide. Higher punishment could effectively limit the extent of illegal use and collect. For the existing feature of bigdata use, it is true that after an overall analysis, it is easy to determine the consumer behavior and use it to make the future decision; however, the current use of data is disordered and random, it is better for the users of data to build a database that help categorize and arrange the data, which can make the future use of data easier and more effective.

## 7. Conclusion

In conclusion, there are more and more implementations of bigdata analysis in consumer behavior and we are still finding a more secure and valid way to do with the implementations. In this document, we first talk about the historical significance of the consumer behavior and the development of bigdata in consumer behavior in recent years. Afterwards, this study discusses the several most common methods and models that used in bigdata analysis. We also give three specific applications, the Decision Tree Regressor model in behaviors in social media, the differential game models in low-carbon tourism and B2B. Based on the overview of the past process of combining bigdata and consumer behavior and several specific applications, we find some limitations including data security problem and data quality problem. In the future, these limitations are going to be fixed by the government supervision and some large database system will be developed to control and use the data more effective. Overall, these results have a clear recognition to the current situation of use of bigdata on consumer behavior theoretically and realistically and have a clear future imagination about the combination of two areas at the same time.

## Author Contribution

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

## References

- [1] Intellipaat. (2013). *Top 10 Big Data Applications in Real Life*. Retrieved from <https://intellipaat.com/blog/10-big-data-examples-application-of-big-data-in-real-life/>
- [2] Jan.vartikao02 (2019). *Big Challenges with Big Data*. Retrieved from Message posted to <https://www.geeksforgeeks.org/big-challenges-with-big-data/>
- [3] Mahmud, M.S., Huang, J.Z., Salloum, S., Emara, T.Z., and Zadatdiynov, K. (2020) *A survey of data partitioning and sampling methods to support big data analysis*. *Big Data Mining and Analytics*, 3(2), 85-101.
- [4] Ahmed, O., Benjelloun, F.Z., Ayoub A.L., Samir B. (2018). *Big Data technologies: A survey*, *Journal of King Saud University-Computer and Information Sciences*, 30(4), 431-448.
- [5] Chaudhary, K., Alam, M., Al-Rakhami, M.S. et al. (2021) *Machine learning-based mathematical modelling for prediction of social media consumer behavior using big data analytics*. *Journal of Big Data* 8, 73.
- [6] Gandomi, A., Haider, M. (2015) *Beyond the hype: Big data concepts, methods, and analytics*. *Int J Inf Manag.*, 35(2): 137-44.
- [7] Bailey, A.A., Bonifield, C.M., Elhai, J.D. (2020) *Modeling consumer engagement on social networking sites: roles of attitudinal and motivational factors*. *J Retail Consumer Serv.* 15, 102348.
- [8] Ma, D., Hu, J., Yao, F. (2021). *Big data empowering low-carbon smart tourism study on low-carbon tourism O2O supply chain considering consumer behaviors and corporate altruistic preferences*, *Computers and Industrial Engineering*, 153, 107061.
- [9] Tang, S., Wang, W., Yan, H., Gang, H. (2015). *Low carbon logistics: Reducing shipment frequency to cut carbon emissions*, *International Journal of Production Economics*, 164, 339-350.
- [10] Christopher, P.H., Sabrina, C., Thornton, P.N (2020). *B2B analytics in the airline market: Harnessing the power of consumer big data*, *Industrial Marketing Management*, 86, 52-64,
- [11] Bormida, M.D. (2021), *The Big Data World: Benefits, Threats and Ethical Challenges*. *Ethical Issues in Covert, Security and Surveillance Research (Advances in Research Ethics and Integrity, Vol. 8)*, Emerald Publishing Limited, Bingley, pp. 71-91.
- [12] Haokun, E., Shuo, H., and Zhang, D. (2019). *Consumer Purchase intention Behavior under Big Data environment*. *Rural Economy and Technology*.
- [13] Liu, J., Fang, L., and Guo, Y. (2018). *The impact of big data use on consumer behavior in the context of mobile commerce*. *The age of business*, 18.