Big Data Usage in Marketing Research

ZHENG Shi[1] JIN Xia[1] ZHENG Wen[2]

*[1]Dalian University for Minorities, School of Economic and Management,
Liaoning Province Dalian Development Zone, 116600, China
[2]Northeastern University in Qinhuangdao, Hebei Province, Qinhuangdao, 066000, China.
E-mail: zhengshi1993@126.com*

Abstract: In the marketing field, the use of big data in research can make us understand consumer deeply. In some areas of market research, big data is already established today. The social media analytics and the use of cookie data to measure internet coverage are two prominent examples. This essay combs through relevant literatures, discusses the big data uses in the marketing research and its contribution for decision-making. It presents a revision of main concepts about marketing research, the new possibilities of use and a reflection about limitations of big data in the marketing research.

Keywords: Big Data, Marketing Research, Marketing Decision Making, user behavior track, data collection, behavior analysis

1. INTRODUCTION

The key of management lies in decision-making. The key of decision-making lies in information. The more correct the information is, the stronger the relevance is, and the greater the probability of success of enterprise decision-making is. At present, the offline traditional research has the problems of great difficulty and high cost, while the network access fixed sample group research and big data research have the problems of insufficient sample representation and difficult statistical inference. Although a large number of literatures are discussing how to make full use of the outstanding advantages of big data, accurately and efficiently reach the target samples, more comprehensively and profoundly mine user characteristics, gain insight into research content, and

complete survey projects with high efficiency and quality; However, there is still a lack of comprehensive literature to analyze and summarize the application of big data in marketing research from the perspective of the whole process of research. This paper combs the relevant research of scholars and institutions in this field and related fields, systematically refines and summarizes the views of experts and scholars on the application of big data in the specific field of marketing research.

The purpose of this paper is to discuss the application of big data in marketing research, put forward the difference between it and the "small data" obtained by traditional research, reflect on its limitations and point out the future trend by analyzing its application status and problems in the marketing research process. For this purpose, the views of researchers and experts will be discussed according to academic publications, and analyzed and faced. Therefore, we can infer the conclusion on this topic.

2. MARKETING RESEARCH DATA IN MARKETING DECISION MAKING

Data in Analyzing and evaluating opportunities

The success of enterprises in the rapidly changing market largely benefits from the efforts in consumer research. When the opportunities and situations related to the use of products can be found, the purpose of research will be achieved. The statistical data of the statistical department and the data summary of relevant industry associations (alliances), and the first-hand data obtained by the research team through field research are the main data sources of market opportunities and evaluation.

Data in Segmentation and Targeting

In order to succeed, enterprises need to find identifiable, measurable, substantive, accessible, responsive and feasible groups. The segmented data comes from both statistical data and information obtained from continuous research on fixed samples and information obtained from digital research and monitoring tools. Behavioral segmentation tends to take advantage of the latter. When the enterprise determines the market segment that the competitors do not recognize or the competitors can not provide sufficient services. Through market analysis, identifying the market to be subdivided, evaluating how to subdivide, and defining micro segmentation strategies, the segmentation strategies are determined [1]. Furthermore, the use of score distribution and

advanced analysis technology can realize the calculation of market dimension and value of life. Consumer behavior and its beliefs and attitudes can be studied through network graphics, which provide relevant information and help enterprises complete the definition of brand and product positioning.

Marketing Research data in Designing and Executing Marketing Plan

Marketing plan is to establish and maintain a feasible adaptive management process between organizational goals, skills, resources and its various changing market opportunities. The formulation and implementation of marketing plans will rely on the following four kinds of research data, (1) product research data. Product development data may benefit from the history of successful products, the analysis of process stages, or the query of creative archives [2]. Both online and offline product surveys are necessary. Retail dive found that 65% of consumers will study the products they want online before entering the store. (2) Price survey data. Offline enterprises will also be strongly affected by the use of online price information. Both online and offline price surveys are necessary. Outer box design points out that 80% of consumers will query product information, comments and prices through smart phones when shopping in physical stores [3]. In April, 2013, a study released by Google Shopper Marketing Council [4] showed that 54% of consumers would use smartphones to compare prices. (3) Distribution research data. Information collection activities that affect the selection of retail outlets and warehouses, together with the operation information of retailers and wholesalers and their response to the marketing policies of manufacturers, are the focus of distribution research. Harvard Business Review research shows that 73% of consumers use multiple channels in the shopping process. Compared with single channel customers, Omni channel customers' in store consumption will increase by 4% and online consumption will increase by 10%. Every time they use more channels, customers will pay more money. (4) Promotional research data. Advertising research and media research, as well as data on prize sales, coupons, sample transactions and research reports on other promotional issues constitute promotional research data.

Monitoring information of marketing performance

Obtaining the monitoring information of brand retail sales and its market share can analyze the marketing performance of enterprises and implement control. In addition, some enterprises will use the data in daily tools such as Google Analytics to evaluate the performance of various marketing channels,

including activities, e-mail marketing, paid advertising, social media marketing, content marketing and website marketing, evaluate the business contribution of marketing, extract original data to build dashboards, analyze personal account databases and contact databases, and analyze paid advertising and social media, Use a common framework to analyze marketing activities, and analyze individual marketing activities at any time.

3. BIG DATA VS. "SMALL DATA"

3.1. Difference Aims :problem-oriented vs. non problem-oriented

The data of market research, which we call "small data" in this article, is collected and analyzed problem oriented. Whether it's qualitative market visits or quantitative data research, we must study with problems. We can't do research for the sake of research, otherwise we can only get a lot of information garbage. The quality of data collected in market research can be controlled by controlling sampling error and non sampling error. The relevant variables of measurement are optimized for specific topics, which can be very extensive. By controlling the sampling error, the representativeness of samples can be guaranteed.

Big data is non problem oriented, and Volker Bosch also pointed out that big data belongs to passive measurement [5]. They are basically domain data, such as UGC data, device data, log data [6]. Because information is defined as the lack of uncertainty in statistical analysis, big data is mixed in solving specific problems because it does not reduce uncertainty, and some of them lack important variables. Therefore, when specific research problems are raised, it will be difficult to find big data, whether the scope of data topics and research topics are consistent, and whether the terminology and variable classification used by data are suitable for research projects, Therefore, big data is uneconomical for solving specific decision-making problems. This is illustrated by Internet coverage research data based on cookies or network centric networks. Even though almost the entire population is the same as the census, key information is missing, such as socio demographic statistics [4].

3.2. Difference Source

The data of market research comes from internal and external data of the enterprise, as well as field research data. There are five basic sources of second-hand data: Book and journal publishers, government sources, mass media sources, industry associations sources, and commercial sources. There is

some overlap between small data and big data. Big data expands the sources of primary and secondary data, such as mass media, commercial sources, etc. Personal data (data collected from fidelity program or email), browsing data collected through cookies and outsourcing data collected from financial institutions, census and credit cards, UGC data, device data, mobile phone data in log data, web search data, online marketing data, etc. all enrich the existing data forms, including both text information and image information [5]. The log information of online marketing activities (i.e. transaction time, price, volume, etc.) is mainly collected from relevant online market platforms (take Taobao [16] and Amazon as examples [7]) through the web crawler tool [7]. AIS data in the market forecast is collected and provided by airborne transceivers, ground and satellite base stations [9]. Big data of industrial manufacturing is identified as having three sources: structured data, unstructured data and semi-structured data. Market data, consumption data and sensor data are classified as structured data; Human interaction data and social media data are considered as unstructured data; Web pages and industrial Internet of things data are classified as semi-structured data [6].

3.3. Difference Design for Deepening insight

Traditional research includes questionnaire design and big data research includes algorithm design. This is the soul of research. Its essence is to deepen insight through data validation.

The traditional data research mainly adopts the method of sampling survey, such as the traditional questionnaire survey. In first tier cities such as Beijing and Shanghai, the effective sample size is more than 300, which is basically considered to represent the whole market. But in the era of big data, in theory, all users can be scanned. The increase of sample size will make the results more accurate.

In recent years, China's top domestic entertainment traffic has been occupied by small fresh meat. From Weibo topics, popularity, to the number of fans, to the initiative of the rice circle, they firmly occupy the commanding height of the industry; This has led to a phenomenon that a large number of female products, such as cosmetics and luxury bags, are endorsed by male celebrities, especially those multinational brands. The result is that most of them are not satisfactory. Since 2022, brands have changed back to female spokesmen. This is a typical case of lacking algorithm design and being brought into the ditch by data; The fans of traffic male stars have higher viscosity and are more willing to make lists for their idols, so they can crush female stars in data. However,

the choice of spokesperson is not only a dimension of traffic data, but also a consideration of brand and product matching, relevant associations, public influence and other aspects; If we only consider the flow data, there will be no correct research conclusion.

First insight, then data; The data reflects the appearance of phenomena and cannot directly get insight, but it can corroborate and correct the insight to a certain extent.

4. USE OF BIG DATA IN MARKETING RESEARCH

Big data technology makes the information obtained from marketing research more extensive. In the long run, big data will change marketing research [5]. The current marketing research related to big data can be roughly divided into three stages: the first stage: the data mainly comes from the OLTP data of the main business, including users, commodity information, transaction data and collection data. The second stage: with the emergence of personalized recommendation system and advertising system, log data has become another data source, and some data can be obtained through cookies. The third stage: with the emergence of apps, the behavior information of wireless users cannot be obtained through traditional logging; Therefore, a number of SDKs developed by enterprises have emerged, through which wireless information can be directly collected. For example, enterprises such as Umeng, talkingdata and Taobao have similar interface programs.

The purpose of the market research is to find the trend and track of the market. Trends of the market include industrial policies, scale, industry growth, profitability, consumption trends, technological changes, upstream changes, channel changes, category development, major trends of head brands, etc; Through the inductive analysis of these data, the overall trend judgment is finally obtained. User behavior trajectory will be described in detail later.

For digital marketing in the era of big data, the first thing to do is to "clean" and mine the data, refine and analyze it in depth, and finally form the feasibility report data for decision-making or prediction. Big data is non problem oriented. In marketing research, it has two main research directions: consumer insight and prediction. As GfK research found, big data lacks important variables in most marketing research; In addition, due to the large number of cases in the big data on consumer behavior, it is still inconclusive to study whether the big data can provide the measurement of consumers and groups with such behavior at the same time [6].

4.1. Use of Big Data in market research

4.1.1. Analysis of user behavior and characteristics

In the era of big data, in theory, all users can be scanned. The increase of sample size will make the results more accurate. Through the tracking of big data, we can understand customers in the whole process of customer problem identification, search, evaluation, decision-making, and after-sales [7-14]. According to what users have done in the past, we can infer what they will do in the future, which is the analysis of user behavior and characteristics, or user behavior trajectory analysis.

User behavior trajectory is composed of static and dynamic parameters. The static parameters in the field of consumer goods are mainly population parameters, age, residence, gender, income, educational background, family status, occupation, belief, social class, etc; The static parameters of enterprise users are industry, scale, income, business characteristics, organizational process, reputation, etc; The platform likes to say "user portrait", which is this part [17].

The dynamic parameters of behavior trajectory are obtained around consumption behavior, such as user identification problems [15], search information, how to make decisions, purchase frequency, purchase quantity, etc. The behavior trajectory can abstract the labeled user model.

Current research on user behavior trajectory focuses on user web behavior trajectory [19], mobile user behavior trajectory [20], user trajectory recognition model [21], user trajectory recognition method and user trajectory recognition system [15][22].

Clickstream is an open source project based on Java, which is mainly used to track the browsing history of users on the web server. Through the historical data, we can analyze the path, bottleneck, hotspot connection and so on. The collection of user behavior trace information is mainly from the page and background, based on the open source clickstream user behavior collection technology, combined with JavaScript script plug-ins, log interception and other methods [23].

For the grasp of product trajectory, it is necessary to master the purchase, sale and inventory of dealers, second batches and terminals, so as to track each product in detail.

Product trajectory refers to the trajectory of products from leaving the factory to circulation, to being purchased and used; The most difficult and important thing to master is the psychological trajectory, which is also composed of static parameters and dynamic parameters.

Static parameters refer to users' personality characteristics, behavior preferences, consumption preferences, values tendencies, and dynamic parameters refer to demand motivation, attitude, cognitive characteristics, risk awareness, sense of experience, expectation, and satisfaction.

The collection of user behavior trajectory information is achieved through JS code insertion, HTTP access interception, java code dynamic implantation and other technologies. Therefore, the implementation of user behavior trajectory requires the cooperation and transformation of the original system with a certain amount of work [23].

Data mining tool is an effective method of analyzing and processing data to access all data sites for multidimensional analysis and decision support [24].

UGC data, web data and transaction data of various mobile applications provide big data sources for data mining. Now, various data mining technologies have been applied to extract user behavior and feature knowledge from UGC data (for text, common tokenization tools and stem analysis technology are used; for photos, gray processing and GLCM method are used.

4.1.2. Customer screening and precision marketing analysis

Big data analysis enhances customer service and product quality for enterprises [25]. Big data analysis is also an intangible resource. It is no longer based on experience and intuition to guide work, but an ability to develop, explore, accumulate, share, and transform information and knowledge to guide decision-making [26]. What many entrepreneurs struggle with is that they can distinguish the most valuable users among the users, friends and fans of the enterprise. They can find countless pieces of information through big data, and use some rules to associate and synthesize them, which can help enterprises screen key target users.

Based on the detailed log information of online marketing (about consumers [27], transaction time [8], products [7], transaction price [7], trade volume [8], etc.), we can deduce new knowledge about customers' marketing behavior, preferences and interests [8], which seems to be a powerful predictor in relevant predictions. Effective big data can also introduce analysis technology to capture other interesting features, such as the potential aspects of text. Rating analysis is used to estimate personal opinions and the importance of each aspect [28]. Another example is semantic based image retrieval to mine information data hidden in p hoto semantic annotation [29].

As an important task of text analysis, text classification has been widely studied, and many methods have emerged, such as text classification

method based on hidden document body generation model of late Dirichlet allocation (LDA), text classification method based on bag of word model and text classification method based on support vector machine (SVM). Research on customer satisfaction through text vector selection of text data (using popular LDA and emotion analysis tools) has made some progress [30][31]; Some scholars also extract useful information through information feature extraction of massive features (using PCA and cluster analysis) [32] [33][34] for user analysis, behavior prediction, customer classification, customer loyalty analysis, etc;

Customer behavior in marketing [35] (for example, audio data and video data from Netflix, youtube, vine and Facebook) are also used in market forecasting. In data analysis, dealing with such data types will only slightly increase the complexity. In the prediction analysis, AI models (especially support vector regression (SVR), CNN and long-term and short-term memory (LSTM)) dominate the prediction models using such UGC data [36][37][38].

4.1.3. *Effectiveness of brand communication and analysis of competitors*

The effectiveness of brand communication can be found through big data analysis. For example, communication trend analysis, content feature analysis, interactive user analysis, positive and negative emotion classification, word-of-mouth category, analysis, product attribute distribution, etc. can be carried out. The communication situation of competitors can be mastered through monitoring, and the industry benchmark user planning can be referred to. According to the user voice planning content, the operation effect of microblog matrix can even be evaluated.

Many enterprises want to know about the dynamics of their competitors. Even though it is difficult to obtain the information of their competitors, they can analyze it through big data monitoring. Collect product review data from various Internet platforms such as shopping websites, social media and product communities, e.g. Qiu Haibin establishes the co-occurrence matrix of target enterprise products and competitor products from massive data through consumer UGC data, with the help of word frequency co-occurrence technology, carries out visual analysis with the help of social network technology, obtains the specific competitive products of the target enterprise, and then statistically calculates the product sales volume, comment emotional value, product network popularity and price of target products and competitive products, and establishes the regression model of product sales volume, With the help of TensorFlow neural network framework developed by Google, a

neural network model for predicting product sales is established [39], and the product competitor identification and sales volume prediction are completed.

4.2. Price trend forecast

Predictive analysis is the core function of big data. Big data prediction is based on big data and prediction model to predict the probability of something in the future. The biggest difference between big data and traditional data analysis is to make the analysis change from "facing the past that has happened" to "facing the future that is about to happen". CPI represents the price fluctuation that has occurred, and big data may help people understand the future price trend and predict inflation or economic crisis in advance. The design of text mining technology for CPI prediction [40], the use of Internet search volume to predict CPI [41], and web crawler technology to promote the integrated development of big data and CPI survey [42], are all applications of big data in price trend prediction.

In terms of sample time, the application of big data can obtain the price and quantity of goods in real time, and eliminate statistical errors by dynamically calculating the CPI index. In terms of sample range, the application of big data makes the collected samples very similar to the population, and the range of samples has been significantly expanded. After the application of big data, it effectively improves the speed of CPI weight adjustment, further improves the transparency of CPI weight adjustment, and highlights the impact of big data on CPI weight [43].

5. LIMITATIONS

5.1. Decision-making is still the responsibility of enterprise decision makers in the Big data era

Big data improves the efficiency of marketing research, but research is still only responsible for analysis and suggestions, and decision-making is still the responsibility of enterprise decision makers.

The purpose of the research is to find the trend and trajectory of the market. It is to assist decision-making, and it is no exception in the era of big data. Big data cannot form decisions independently, and there will be many cases of improper application of big data in the future. Like apple, it will bear the huge cost of third-party research data support that despises independent systems.

The automation of decision-making realized by the prediction model [5] also represents risk, because no matter how good the prediction model is, the

model is still a binary method to understand the finite theoretical situation. At least for now, the analytical model will be responsible for the implementation of analysis and recommendations, but decision-making is still the responsibility of mankind.

5.2.   Big data is not accurate, but mixed, and needs to be used correctly

In the era of small samples, data is scarce, so we first pursue the accuracy of data, and then the accuracy of results. However, in the era of big data, data is too complex to be accurate. Some scholars pointed out that in view of the low information density of big data, big data is a lean mine, and the output ratio is not necessarily good. In the era of big data, the hybridity of data becomes inevitable: first, as the amount of data becomes larger, the error rate will also increase. Second, data comes from various systems, which increases the degree of promiscuity. Third, data formats are various, and data cleaning is very difficult. In addition, there are data shortages and discontinuities due to big data. Therefore, big data (almost always) is not representative; Big data is (almost always) flawed; Big data (almost always) lacks important variables; These conclusions have almost become an academic consensus [5].

It is estimated that only 5% of the data is structured and can be stored in traditional databases. If mixing is not accepted, the remaining 95% of unstructured data cannot be used, such as web pages and audio and video resources [5]. By accepting uncertainty, we open a window to a world that has not been touched. Big data requires us to change. We must be able to accept confusion and uncertainty. Accuracy seems to have always been the support of our life, but the idea that there is only one answer to each question is untenable. Once we acknowledge this fact and even support it, we will be one step closer to the truth.

In the era of small data, we need to use the least data to get the most information. The essence of big data is not the absolute value of the amount of data, but the method of using all available data for analysis instead of random sampling analysis. In this sense, big data emphasizes not data, but methodology.

Big data can make up for the deficiencies of traditional research, but big data research can't do what traditional research can't do. One of the most prominent is the grasp of users' psychological path. Without the analysis logic of professional research, without the support of data analysis technology and huge equipment, the original big data is just a huge "waste". Therefore, an important aspect of big data application is the correct use of information [44].

5.3. Big data still lacks clear decision guidance

Big data implements value orientation rather than decision-making demand orientation, that is, mining in massive data to see what valuable information can be found. In a macroeconomic or social sense, there is no problem at all. However, from the perspective of enterprise micro decision-making, big data with large amount of data, great noise and low value density may be difficult to meet the applicability requirements of enterprise decision-making for specific data value.

There is no right to speak without investigation, but how much say research should have in the final decision-making has always been controversial; The general understanding is that research is to collect information, and insight is to process information, which can break the essence; Insight without research is blind and blind; Research without insight is mindless and formalistic. In the information age, it is easier for enterprises to obtain information and people's behavior trajectories to be mastered; This makes it more important for enterprises to obtain more accurate and effective information and establish market intelligence systems for enterprises through big data. Big data intelligence replaces some of the previous ways of obtaining information, such as reading media reports, such as inquiring about various gossip; But big data cannot replace insight, let alone make decisions automatic; Research and insight still perform their respective duties, make insights into big data information, and use big data information to verify and deepen insight conclusions; This is the essential relationship between the two and will not change. The value of big data to enterprise marketing is not only reflected in the information system, but also plays an important role in sales management, promotion management, product and brand management, supply chain management, but the basis is the information system.

How to deal with these ultra-high dimensional data, mine its potential value, and develop a data flow model suitable for the new manufacturing environment is a challenging problem [45]. With the further development of data storage and analysis technology, big data-driven analysis is an important driving force to create the main value of manufacturing [46].

6. FINAL CONSIDERATIONS

As discussed above, due to the many characteristics of big data, the ultimate problem in the application of big data in market research is: who should have the right to collect big data, who should have the right to use big data, and who should have the right to open big data. As Nunan and Domenico [50] put

forward, autonomy through ESOMAR (European Society for public opinion and Market Research) [51] is another way to combat abuse and excessive behavior, And make good use of information. Nunan and Domenico [50] proposed that the current esomar[51] rules include the right to be forgotten (which may require deleting the history), the right to expire the data (which may supplement the forgotten right, and the transaction data may also expire), and the ownership of the social graph (individuals should know the collected information). In 2016, a number of manufacturers cooperating with Talkingdata found that the products released on Google play were removed from the shelves, and the instructions for the removal said in the email, "in violation of the developer terms" and pointed out that it was caused by the problems of talkingdata SDK. Talkingdata responded to the SDK, causing it to be removed from the shelf: google play reviewed and adjusted. This may be a typical case of privacy infringement.

In addition, big data must be scientifically evaluated in market research. Passive measurement is the efficiency of big data in market research. It created what was once a dream, that is, economies of scale. Big data can be used like a microscope to see the fuzzy structure, even completely unrecognizable through traditional market research [5]. However, big data is not omnipotent. It is flawed. There must be appropriate personnel in the field of application and research after evaluation. At this level, abuse and excessive behavior should also be avoided. According to gobble (2013) [2], the McKinsey report identified 190000 qualified personnel working in data analysis related positions today. In addition to those directly related to the analysis, don Schulte [47] still emphasizes the need for people with "real-life" experience to be able to interpret the information generated by the algorithm. "If the basic understanding of customers does not exist, it is built into the analysis model, and it does not matter how many iterations or how fast the data goes through. This output is worthless. Therefore, the construction of enterprise public information disclosure system [48][49] may be a promising way to promote the application of UGC data in prediction.

References

[1]   Cravens, D.W. and Piercy, N.F. (2008). Marketing estratégico. 8th Edition, McGraw Hill, São Paulo.

[2]   Gobble, M.M. (2013). Big Data: The Next Big Thing in Innovation. Research-Technology Management, 56, 64-67.http://dx.doi. org/10.5437/08956308X5601005

[3]   https://www.sohu.com/a/314506582_700450

[4] Google Shopper Marketing Agency Council (2013). Mobile In-Store Research: How Is Store Shoppers Are Using Mobile Devices, 37. http://www.marcresearch. com/pdf/Mobile_InStore_Research_Study.pdf

[5] Volker Bosch. Big Data in Market Research: Why More Data Does Not Automatically Mean Better Information. GfK MIR / / GfK Research. Vol. 8, No. 2, 2016,56-63

[6] Chunquan Li , Yaqiong Chen, Yuling Shang.A review of industrial big data for decision making in intelligent manufacturing. *Engineering Science and Technology, an International Journal.*2022,29:1-16

[7] S. Carta, A. Medda, A. Pili, D.R. Recupero, R. Saia, Forecasting e-commerce products prices by combining an autoregressive integrated moving average (ARIMA) model and Google trends data, Future Internet 11 (1) (2019) 5,https:// doi .org /10 .3390 /fi11010005.

[8] J. Zhang, A. Simeone, P. Gu, B. Hong, Product features characterization and customers' preferences prediction based on purchasing data, CIRP Ann. 67 (1) (2018) 149–152, https://doi .org /10 .1016 /j .cirp .2018 .04 .020.

[9] M. Gao, G. Shi, S. Li, Online prediction of ship behavior with automatic identification system sensor data using bidirectional long short-term memory recurrent neural network, Sensors 18 (12) (2018) 4211, https://doi .org /10 .3390 /s18124211.

[10] M. Inthachot, V. Boonjing, S. Intakosum, Artificial neural network and genetic algorithm hybrid intelligence for predicting Thai stock price index trend, Comput. Intell. Neurosci. 15 (2016) 1–8, https://doi .org /10 .1155 /2016 /3045254.

[11] A. Goyal, S. Krishnamurthy, S. Kulkarni, R. Kumar, M. Vartak, M.A. Lanham, A solution to forecast demand using long short-term memory recurrent neural networks for time series forecasting, in: Midwest Decision Sciences Institute Conference, IEEE, 2018, pp. 1–18.

[12] J. Wang, R. Hou, C. Wang, L. Shen, Improved v-support vector regression model based on variable selection and brain storm optimization for stock price forecasting, Appl. Soft Comput. J. 49 (2016) 164–178, https://doi .org /10.1016 /j.asoc .2016 .07.024.

[13] F. Yoseph, M. Heikkilä, D. Howard, Outliers identification model in point-of-sales data using enhanced normal distribution method, in: 2019 *International Conference on Machine Learning and Data Engineering, IEEE,* 2019, pp. 72–78.

[14] A. Goyal, S. Krishnamurthy, S. Kulkarni, R. Kumar, M. Vartak, M.A. Lanham, A solution to forecast demand using long short-term memory recurrent neural networks for time series forecasting, in: Midwest Decision Sciences Institute Conference, *IEEE,* 2018, pp. 1–18.

[15] Meng Wei;Dong Kai. Research on user behavior track and early warning system based on big data. *Cyberspace Security.* 2019,10(12): 98-102.

[16] O. Schaer, N. Kourentzes, R. Fildes, Demand forecasting with user-generated online information, Int. J. Forecast. 35 (1) (2019) 197–212, https://doi .org /10 .1016 /j .ijforecast .2018 .03 .005.

[17] KRIEGEL H. Data Science/Data Mining[J].Digital Welt, 2019,3(1):7-8.

[18] YuSong. Mining Association Rules in the Big Data of Guizhou Cigarette Brand (M), 2017) https://kns.cnki.net/KCMS/detail/detail.aspx?dbname= CMFD202001&filename=1017195571.nh

[19] Kori S, Zhu Y, Yamaguchi K, et al. Analysis of user's behaviour based on search intentions for information retrieval using search engines[C]// 2015 Conference on Technologies and Applications of Artificial Intelligence (TAAI). IEEE, 2015.

[20] WANG Su-fang. Analysis of mobile user behavior in big data environment. Information Technology and Informatization, 2020,11:250-252

[21] ZHANG WeiLI YangZHANG JiWANG Jian-Yong. A User Trajectory Identification Model with Fusion of Spatio-Temporal Behavior and Social Relation. *Chinese Journal of Computers.* 2021,44(11):2173-2188.

[22] Wang Yunan. User Behavior Analysis Method Research Based on Trajectory Data. Shenyang Ligong University, 2020.DOI:10.27323/d.cnki.gsgyc.2020.000462.

[23] LinXueyun. Research on User Behavior Trajectory. *Journal of Chongqing Industrial and Commercial University* (Natural Science Edition).2015,32(01):78-82. DOI:10.16055/j.issn.1672-058X.2015.0001.019.

[24] K. Gandhi, B. Schmidt, A.H. Ng, Towards data mining based decision support in manufacturing maintenance, in: Proc. CIRP Conf., 2018, pp. 261–265.

[25] R.H. Hamilton, W.A. Sodeman, The questions we ask: Opportunities and challenges for using big data analytics to strategically manage human capital resources, Bus. Horiz. 63 (2020) 85–95, https://doi.org/10.1016/j. bushor.2019.10.001

[26] J. Lee, J.H. Lee, Constructing efficient regional hazardous weather prediction models through big data analysis, J. Intell. Fuzzy Syst. 16 (1) (2016) 1–12,https:// doi .org /10 .5391 /ijfis .2016 .16 .1.1.

[27] K. Fang, Y. Jiang, M. Song, Customer profitability forecasting using big data analytics: a case study of the insurance industry, Comput. Ind. Eng. 101 (2016)554–564, https://doi .org /10 .1016 /j .cie .2016 .09 .011.

[28] H. Wang, Y. Lu, C. Zhai, Latent aspect rating analysis on review text data: a rating regression approach, in: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2010, pp. 783–792.

[29] K. Song, F. Li, F. Long, J. Wang, Q. Ling, Discriminative deep feature learning for semantic-based image retrieval, IEEE Access 6 (2018) 44268–44280, https:// doi .org /10.1109 ACCESS .2018 .2862464.

[30] DuChangShun. Research on Key Technologies of Public Opinion Sentiment Analysis for Segmented Fields. Beijing Jiaotong University. 2019.

[31] LI Rui;ZHANG Wei-bin. Application of Data Mining Technology Based on TF-IDF Algorithm and LDA Topic Model in Power Customer Complaint Text. *Techniques of Automation and Applications,* 2018,37(11): 46-50.

[32] Dengyingfan. Research on user analysis and behavior prediction driven by big data [D]. Beijing University of Posts and Telecommunications,2020. DOI:10.26969/d.cnki.gbydu.2020.002814.

[33] Huangpuhancong, Xiaozhaodi, Yuyong Zhong. Power consumer value classification and application based on entropy weight method and improved PCA clustering algorithm [J]. Modern Electronics Technique,2017,40(07):183-186. DOI:10.16652/j.issn.1004-373x.2017.07.048.

[34] ZhuBangZhu. Evaluation and Improvement of Customer Loyalty Based on the PCA/DEA Model [J]. Journal of Wuhan University of Technology (information and Management Engineering Edition),2008(01):109-113.

[35] S. Tanuwijaya, A. Alamsyah, M. Ariyanti, Mobile customer behaviour predictive analysis for targeting Netflix potential customer, in: 2021 9th International Conference on Information and Communication Technology, ICoICT,IEEE, 2021, pp. 348–352.

[36] T. Trzci´ nski, P. Rokita, Predicting popularity of online videos using support vector regression, IEEE Trans. Multimed. 19 (2017) 2561–2570, https://doi .org /10 .1109 /TMM .2017.2695439.

[37] W. Liu, Z. Duanmu, Z. Wang, End-to-end blind quality assessment of compressed videos using deep neural networks, in: ACM Multimedia, 2018, pp. 546–554.

[38] H. Dou, W.X. Zhao, Y. Zhao, D. Dong, J.-R. Wen, E.Y. Chang, Predicting the popularity of online content with knowledge-enhanced neural networks, in:ACM KDD, 2018.

[39] QiuHaiBin. Research on product competitor identification and sales volume prediction based on consumer UGC[D]. Jilin University, 2019.

[40] Tang Xiaobin;Dong Manru;Xu Rong. Design and Application of Text Mining Technology for CPI Prediction Based on Big Data [J]. Statistical Research,2021,38(08):146-160.DOI:10.19343/j.cnki.11-1302/c.2021.08.012.

[41] Ouyang Mengqian;Zhou Xianbo;Zhu Junmei. Apply Internet Search Index to Predict CPI in the Big Data Era——Based on Joint Application

of LASSO and Kernel Partial Least Squares [J]. Quarterly Journal of Finance,2020,14(02):112-136.

[42] Hujianwei,Qiuhaiyan. Web crawler technology promotes the integrated development of big data and CPI investigation [N]. China Information Daily,2019-09-12(004).DOI:10.38309/n.cnki.nzgxx.2019.001360.

[43] LiAng. Research on the influence of big data on CPI statistics and method improvement [J]. China Management Informationization, 2019,22(02):193-194.

[44] Alexandre Borba Salvador, Ana Akemi Ikeda. Big Data Usage in the Marketing Information System. Journal of Data Analysis and Information Processing, 2014, 2, 77-85

[45] S. Kumar, K.K. Mohbey, A review on big data based parallel and distributed approaches of pattern mining, J. King. Saud. Univ. Comput. Inf. Sci. doi:10.1016/j.jksuci.2019.09.006.

[46] T. Ahmad, H. Chen, W.A. Shah, Effective bulk energy consumption control and management for power utilities using artificial intelligence techniques under conventional and renewable energy resources, Int. J. Electr. Power Energy Syst.109 (2019) 242–258, https://doi .org /10 .1016 /j .ijepes .2019 .02 .023.

[47] Schultz, D. (2012) Can Big Data Do It All ? Marketing News, November, 9.

[48] Ling Tang , Jieyi Li , Hongchuan Du , Ling Li, Jun Wu , Shouyang Wang. Big Data in Forecasting Research: A Literature Review. Big Data Research.2022,27:1-30

[49] J. Wang, B. Zhang, Quality of environmental information disclosure and enterprise characteristics, Manag. Environ. Qual. An Int. J. 30 (2019) 963–979,https://doi .org /10 .1108 /MEQ -11 -2018 -0194.

[50] Nunan, D. and Domenico, M.Di. (2013) Market Research and the Ethics of Big Data Market Research and the Ethics of Big Data. International Journal of Market Research, 55, 2-13.

[51] ESOMAR. http://www.esomar.org/utilities/news-multimedia/video. php?idvideo=57