

UDC: 658.8

JEL: L81

CUSTOMER SATISFACTION FACTORS IN ONLINE RETAIL: ONLINE REVIEW ANALYSIS

M. Haddadi, V. A. Rebiazina

HSE University,
20, ul. Myasnitskaya, Moscow, 101000, Russian Federation

For citation: Haddadi M., Rebiazina V. A. 2023. Customer satisfaction factors in online retail: Online review analysis. *Vestnik of Saint Petersburg University. Management* 22 (1): 3–22.
<http://doi.org/10.21638/11701/spbu08.2023.101>

This paper aims to analyze online customers' reviews and gain insight into customer satisfaction factors drawn from a big user-generated content of US online retailers. This user-generated content has become more insightful, especially in the COVID-19 era as the result of a rapid increase in online shopping. This study uses a big textual data of 5 340 786 online reviews which was collected from the platform bizratesurvey.com. The study focuses on individual customers' reviews of 839 US online retailers. Word frequency analysis and latent dirichlet allocation methods were used to process the data. The results revealed three main topics related to the ease of use, product, and delivery that were mentioned by highly satisfied customers of US online retailers. The authors have labeled the topics properly by considering at least twenty of the most probable words in each topic. The results provide a pathway for online retail executives for enhancing shopping experiences through the ease of purchasing, the improvement of product quality and delivery for customers. Practitioners can replicate the process of data analysis fulfilled in this study in order to monitor customer feedback. The findings also provide a new way of using big textual data from customer reviews in further studies.

Keywords: COVID-19, customer satisfaction, online retail, text analytics, online review.

INTRODUCTION

The retail industry has experienced major disruptions because of the COVID-19 pandemic and businesses need to adapt to the situation and react immediately [Pantano et al., 2020]. In the COVID-19 era, online shopping has grown to tremendous magnitude, and new online shoppers and existing customers are eager to purchase goods in af-

This research has been conducted within the fundamental research project “Digitalization as a driving force of open innovation and co-creation: Implications for value creation and value capture” as a part of the HSE Graduate School of Business Research Program in 2021–2023 (Protocol No.23 dated 22.06.2021 of the HSE GSB Research Committee).

© St. Petersburg State University, 2023

fordable ways. Thus, many companies moved online to increase their revenues [Barnes, 2020]. This movement, to adapt e-commerce by new online shoppers, has been evidenced in many countries in the pandemic era [Oliveira et al., 2021]. Although, in some parts of the world, a few obstacles have arisen for newcomers when shopping online, such as prolonged delivery time and lengthy product searching processes [Prasad, Srivastava, 2021].

User-generated content, such as online customer reviews in the COVID-19 era, can assist newcomers and existing customers to enhance their online shopping experience. Researchers have identified the impact of online textual reviews on customer decision-making and product sales [Zhang et al., 2014; Hernández-Méndez, Muñoz-Leiva, Sánchez-Fernández, 2015]. These types of unstructured textual data can guide online business owners to recognize the advantages or disadvantages of their services or products compared to their competitors, through the lens of consumers. There are only a few studies that identify customer satisfaction factors for online shopping experience, especially in the COVID-19 era, from Big textual data. In addition, many past studies, conducted on customer satisfaction in online shopping, have attempted to identify the influencing factors using the scale measurement method. Thus, the authors aim to address this research gap by using topic modeling as an unsupervised machine learning technique to analyze online customers' reviews in order to understand the factors that influence customer satisfaction in online shopping. The main theoretical foundation of this study is signaling theory [Spence, 1973] and expectation-disconfirmation theory [Oliver, 1977; 1980]. The authors for this study utilize large text data consisting of 5 340 786 online reviews written by individual customers including experienced ones or new customers about their online shopping experiences across 839 US online retailers.

To address the aforementioned literature gap in order to assist online retailers to better meet customers' needs during the COVID-19 era, the following research question was formulated to guide this study: *which factors contribute to customer satisfaction in online shopping?*

This study makes two main contributions. Firstly, it contributes to an original insight into a high volume of user-generated content, in the form of online reviews using a topic modeling approach about the factors affecting customer satisfaction in the online retail sector during the COVID-19 era. The second contribution is the use of real business data to analyze customer satisfaction in online retail, which can be more rigorous compared with the traditional measurement scale approaches in the marketing field. Existing research has used the traditional measurement scale (e.g., [Chang, Chen, 2009; Kim, Jin, Swinney, 2009; Reibstein, 2002; Szymanski, Hise, 2000]). Thus, utilizing large volumes of customer-generated online reviews and text analytics techniques in this study provide an opportunity to gain more in-depth perspectives by using a broader and more representative set of customer-relevant keywords. In summary, the authors seek to draw their contributions with specific findings across the literature related to the analysis of customer textual reviews and customer satisfaction in the online retail context.

The structure of this study is as follows. In the first section, the authors provided a theoretical background on the impact of COVID-19 in the online retailing industry, as well as the relevance of online reviews, and finally, customer satisfaction factors of online shopping. The second section describes the data and methods of analysis. The third section presents the result of the study. Finally, the authors conclude the discussion and suggest a future research direction.

THEORETICAL FRAMEWORK

This research seeks to explore customer satisfaction factors in the context of online retail through analyzing online reviews. In this regard, this section firstly describes the literature on the impact of the COVID-19 pandemic on the online retail industry. Second, it discusses the literature on the relevance of online reviews in the current digital age. Finally, customer satisfaction factors are explained in online shopping.

The COVID-19 impact on the online retail industry. In the COVID-19 era, the absence of perceived control levels is positively and significantly associated with shopping. A number of studies concerning this era indicated new customer behaviors, such as panic buying, to control uncertainty. For instance, panic buying might be caused by a fear of the unknown as purchasing becomes a way to control negative emotions [Yuen et al., 2020; Barnes, Diaz, Arnaboldi, 2021]. However, some other studies have specifically focused on changes in customer behavior in the online context. P. Forster and Y. Tang [Forster, Tang, 2005] found that online shopping has grown because of fears of being infected by the virus. R. Kim [Kim, 2020] is one of the first to investigate the pandemic as a booster of structural changes in consumption and digital market transformation, as well as the long-term impact on consumer behavior towards online retailing. He believes that the pandemic has forced many businesses to operate online and forge the necessary transitions to survive, in order to sell their product or service. R. Kim also suggested that managers need to develop innovative digital sales because of the growth of online shopping in the COVID-19 era. Similar to Kim's findings [Kim, 2020] A. Ghandour and B. Woodford [Ghandour, Woodford, 2020] have used a regression method in their study to analyze online shopping data in the United Arab Emirates and have concluded similarly that the pandemic has a positive effect on consumers' tendency to shop online.

E. Pantano and co-authors argue that online retailers in the COVID-19 era have to deal with a range of challenges that executives need to respond quickly in order to keep their customers satisfied during the pandemic [Pantano et al., 2020]. Consequently, retailers who have failed to adapt to COVID-19 and have not considered any change in their operations are facing a crisis. In a similar vein, a majority of other studies have also concluded that the COVID-19 pandemic has dramatically changed consumer behavior and forced consumers to shop online to acquire their desired product or service [Hashem, 2020; Salem, Nor, 2020].

Relevance of online reviews. Every organization, across various sectors, generates enormous amounts of information and executives face the challenge to utilize the infor-

mation they receive to the best of their ability [LaValle et al., 2011]. On the other hand, customers communicate their opinions on their social networks through word of mouth. Word of mouth is defined as “all informal communications directed at other consumers about the ownership, usage, or characteristics of particular goods and services or their sellers” [Westbrook, 1987, p. 261], is a substantial determinant of consumer behavior [Bansal, Voyer, 2000]. As a form of electronic word of mouth (eWOM), online reviews are peer opinions published on websites. Previous research has concluded that opinions in online product reviews can have a significant impact on consumers’ purchasing decisions [Awad, Ragowsky, 2008]. By using these textual unstructured data, an organization can gain insight into customers’ perceptions of themselves and their competitors [Atul Khedkar, Shinde, 2018]. Online reviews have now evolved into a dimension that can easily be accessed by anyone from anywhere, as well as facilitating and influencing purchasing decisions as customers express their criticism when they buy a service or product [Ho-Dac, Carson, Moore, 2013].

S.-H. Hsu argues that customers tend to discontinue their purchases from a company when they feel dissatisfied with the service or product they bought [Hsu, 2008]. In this situation, dissatisfied customers spread negative information to potential customers, mainly in the form of user-generated content. Author also mentioned that negative shopping experiences that a customer had with a particular retailer can easily and quickly spread to potential new customers in the online context through various platforms that offer the option of submitting customer reviews about online retailers. Hence, it has been extremely important in the COVID-19 era to satisfy customers in the online shopping environment because of the rapid growth in online retail. Due to all these reasons, analyzing customers’ online reviews to understand their thoughts, perceptions and feelings using new methods such as text analytics may be a necessary need for a business.

Customer satisfaction factors. The underlying theoretical foundation of this research to determine factors of customer satisfaction in the online context is mainly based on two theories: signaling theory and expectation-disconfirmation theory. Signaling theory argues that information can be used as a signal for customers to have expectations about features of products and services [Spence, 1973]. The authors in this study considered online reviews as a type of information act like a signal. Another theoretical basis for this study is expectation disconfirmation theory. This theory states that expectations combined with perceived performance lead to post-purchase satisfaction. This effect is mediated by a positive or negative discrepancy between expectations and performance. If the product/service exceeds expectations, it will lead to post-purchase satisfaction (positive disconfirmation). If the product/service does not meet expectations, the consumer is likely to remain unsatisfied (negative disconfirmation) [Oliver, 1977; 1980]. In this study, customers express their satisfaction or dissatisfaction in their online reviews.

Customer satisfaction is critical to customer retention during the COVID-19 pandemic [Al-Ghraibah, 2020]. Numerous researchers have studied factors affecting customer satisfaction in previous years. Several marketing researchers have studied influ-

encing constructs towards customer satisfaction in the online context with help of different methods. Z. Yang et al. [Yang et al., 2005] found that ease of use is apparently very critical because online transactions are complex and confusing for many customers. This includes ease of navigation, user interface, intuitiveness, and search tools that minimize customer efforts when shopping online. X. Liu with co-authors suggest that eight constructs including product information quality, website design, security, customer service, merchandising, transaction capability, response, payment and delivery are strongly predictive of customer satisfaction in online shopping [Liu et al., 2008]. The analytical results conducted in [Chang, Chen, 2009] demonstrated that customer interface quality and perceived security positively affected customer satisfaction. In the same year, F. Zeng with co-authors [Zeng et al., 2009] stated that product offering, security, customer service, ease of use, fulfillment, and reliability influences customer satisfaction. In a similar vein, authors of study [Kim, Jin, Swinney, 2009] had realized that factors including website design, security, fulfillment and reliability have a significant influence on customer satisfaction.

Y. Vakulenko with co-authors has found a significant impact of last-mile delivery on the relationship between online shopping and overall customer satisfaction [Vakulenko et al., 2019]. The last mile delivery construct is defined as receiving all necessary information about the product delivery and full control of the delivery process by the customer and finally receiving the product on time. In a similar way, U. Tandon and R. Kiran [Tandon, Kiran, 2019] has paid special attention to the factor named “Pay on delivery” (POD) as new constructs that have a significant impact on customer satisfaction. Delivery is also stated in [Al-Jahwari et al., 2018] as influencing factors on customer satisfaction. Later on, P. Deyalage and D. Kulathunga [Deyalage, Kulathunga, 2020] have conducted a systematic literature review study that summarized 51 factors that influence customer satisfaction among 41 studies conducted between 2000 and 2019. They have concluded that during the past two-decades factors including website design, security, customer service, product information quality, convenience and delivery are among most frequent factors towards customer satisfaction among all the factors in their study. Following their study, P. Merugu and V. Mohan [Merugu, Mohan, 2020] identified ease of use, service reliability, responsiveness, assurance, and security as the major determinants of customer satisfaction towards online shopping. In a similar vein, N. Bahari with co-authors has conducted a study in the COVID-19 era and their findings indicated that product quality, security, and shipping significantly affect customer satisfaction in online shopping [Bahari et al., 2021].

METHODOLOGY

This study uses customer-generated content to understand their thoughts, feelings, and perceptions about online shopping experience. Text analytics with big data is rapidly growing in business and marketing research to help overcome several of the biases that are common in traditional survey-based research [Guo, Barnes, Jia, 2017; Berger et al., 2020].

Traditional survey measurement is usually limited to delivering empirical findings using small samples of thousands of observations through self-reported questionnaires whereas user-generated online content can reveal insightful glimpses into people's thoughts, feelings and behavior on a very large scale (hundreds of thousands customer reviews).

This approach helps to avoid some of the biases of the survey instruments [Barnes et al., 2020], and also some common method bias stated by P. Podsakoff with co-authors [Podsakoff et al., 2003] and finally consumer inattention bias [Brosnan, Babakhani, Dolnicar, 2019]. In this section, several text analytics techniques make sense out of a huge volume of online customer reviews. The authors specially opted for latent dirichlet allocation (LDA) as an unsupervised machine learning method to identify main topics that contribute to customer satisfaction in online shopping. Thus, by analyzing customer reviews about online retailers, this research reveals the main factors that contribute to customer satisfaction in the online retail context. The authors first indicate the source of the data and then proceed to data pre-processing and data analysis methods.

Data collection. The secondary data for this article has been collected from bizratesurveys.com. This website is a subdomain of Bizrate Company that focuses on customer insights and analytics. Bizrate has a very high traffic rating among US websites in the Alexa ranking, which increases the external validity of data in this research. Moreover, Bizrate's data is widely used in studies published in leading journals covering customer satisfaction in the online context (e.g., [Reibstein, 2002; Cao, Gruca, Klemz, 2003]). This company analyzes data from online retailers in various industries and gathers feedback from registered customers at two stages of online purchase experience and post-purchase. They have provided a source of customer-oriented feedback regarding online shopping experience that rates online retailers according to their customer satisfaction. At each of these two stages of the survey, customers can submit their reviews as open-ended questions about the product or service they have bought. Most of the customers usually share their whole feedback during the first stage and express their feelings and perceptions about the entire online shopping experience. Few of the customers put their feedback separately at the post purchase survey. In addition, the reviews come from buyers, not sufferers, browsers, or those seeking information. Consequently, the reviews and ratings tend to be positive [Reibstein, 2002]. Figure 1 shows how customer feedback about site experience appears on bizratesurveys.com. Table 1 shows all variables data types that have been collected.

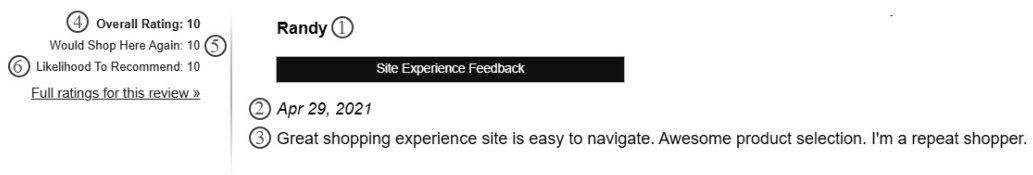


Figure 1. Customer review, April 2021

Based on: Bizrate Surveys. URL: www.bizratesurveys.com (accessed: 25.04.2021).

Table 1. Variables data type

No	Variable	Data type
1	Author	Character
2	Review date	Date
3	Site experience feedback	Character
4	Overall rating	Numeric
5	Would shop here again	Numeric
6	Likelihood to recommend	Numeric

This article's data is a large data set of online reviews (approximately 2.5 GB) and was scraped during approximately 14 days, from 25 April until 9 May 2021. The number of online reviews that were collected in this study from the public web data is approximately 5 340 786. The authors intend to analyze only online reviews of fully satisfied customers from all scraped reviews. The reasons that authors choose only fully satisfied customers is related to the tendency of reviews to be positive on the Bizrate platform [Reibstein, 2002], as well as to avoid any bias in the results of topic modeling. Thus, 3 291 660 reviews of highly satisfied customers who have submitted their reviews about 839 US online retailers have been considered for the data analysis part. These selected reviews are approximately 61% of the entire collected dataset. Selected online reviews submission dates are for the period of the COVID-19 era, from January 3, 2020 until May 9, 2021.

Data preprocessing. In this article, the data processing and preparation process consists of three main steps prior to the application of data analysis methods. All three steps have been done in R Studio. In the first step, the type of data that were scraped from the website was modified. Hence, any variable with a date or number that was stored as a character type has been changed to its correct data type. In this phase, all values that were stored as "null" or "N/A" were replaced by blank values to avoid mistakes in calculating the true number of observations.

In the second step of preparing the data for analysis, only the textual reviews of completely satisfied customers, whose overall satisfaction rating score is the highest value with a numerical value of 10 out of 10 point Likert scale, are selected. Following this, several transformations have been performed on the customers' reviews in order to make the analysis more efficient. These transformations entail making all letters lowercase, removing punctuation marks, removing any numbers, removing extra space between letters, and finally, removing stop words. The authors have an approach to add

several customized words to common stop words. The reason for using this approach is to reduce analysis time by removing unnecessary words before the main parts of the analysis. These new stop words include several common adverbs, past tense verbs, and several nonsense words. For instance, words such as “get”, “always”, “found”, “really”, “like”, “one”, “necessary”, and “years” have been added to the stop words list. In this study, 255 stop words have been considered.

The final step of data preprocessing is tokenization. This step is about splitting each sentence into smaller units called tokens. These tokens can be numbers, n-grams, words and symbols. In this study, tokens are words because of the removal of numbers and punctuation marks in the previous step.

Analysis methods. The authors use several text mining methods to discover valuable information from large collections of unstructured text data [Losiewicz, Oard, Kostoff, 2000]. There is an approach to using a method to get a general understanding of text data. One of the best ways to get this general insight about text data is through word frequency analysis that has been performed on a bag of words in R programming. Following the use of word frequency analysis and a grasp of the most frequent words, the authors intend to use LDA method to understand the main topics of the reviews. A topic is a collection of words that often occur together in customer’ reviews. The authors implement LDA on document-term matrix (DTM) which has been made after tokenization, with the help of the `cast_dtm` function in R programming. “Gibbs sampling” — a randomized search algorithm, which is a type of Monte Carlo Markov Chain algorithm, has been set for the LDA method. Finally, the seed value has been specified to ensure reproducibility of results between runs of LDA function. The alpha parameter, which is the density of the document topic, and the beta parameter, which is the topic word density, have not been set and therefore retained their default values in the LDA models, respectively. The iteration value, which represents the number of Gibbs iterations, has been defined as the default value in all DTM objects.

There are two approaches to determine the value of K in the LDA modeling that define the number of topics. The first approach is to look at the coherence of topics. In this approach, the authors can conclude that coherence is low, while words that typically should not occur together appear in a topic. Another approach entails considering quantitative measures of fit: log-likelihood and perplexity. The log-likelihood is a measure of how plausible the model parameters are, given that the perplexity is the measure of “surprise” when the model receives new data. The perplexity is a positive number, and the fewer surprises the better. Hence, a model with less perplexity is preferable. The authors’ approach to setting the number K in the LDA modeling is based on the most frequent words and the coherence of topics. The authors did not consider the second approach because of the large volume of data. Running an algorithm to fit the model for different numbers of clusters and finding the best possible K out of plots of the likelihood logarithm and perplexity is not an appropriate approach for this study. Concerning the topic labeling that needs to be determined manually [Blei, 2012], the authors labeled topics by obtaining the opinions of two experts in the field in order to reduce subjectivity.

RESULT

This section outlines the result of data analysis for 3 291 660 customer reviews. Highly satisfied customers have submitted these reviews about their purchase experience from 839 US online retailers. Descriptive statistics for the number of reviews for each retailer are shown in Table 2.

Table 2. Descriptive statistic for number of reviews about US online retailers

Min	Quartile 1	Median	Quartile 3	Max	Mean	Standard deviation
2	100	469	2 454	114 740	3 923	10 660

Moreover, Figure 2 shows the number of reviews that have been collected.

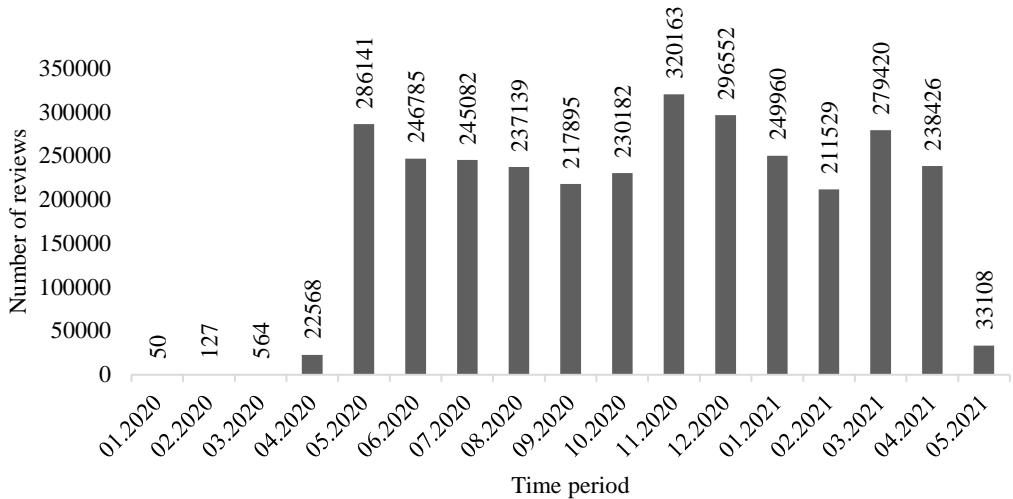


Figure 2. Number of highly satisfied customers' reviews, January 2020 — May 2021

As shown in Figure 2, the number of reviews at the beginning of the period in January, February, March and April 2020 is low, but it varies slightly in other months. The highest number of reviews is for November 2020 with 320 163 reviews.

Word frequency analysis. The result of the word frequency analysis reveals the most frequent terms in customers' reviews. The first word that occurs most prominently in all the reviews of fully satisfied customers about their online shopping experience is "easy". The second word that appears most often in all reviews is "products". The words "experience" and "order" rank third and fourth in terms of frequency of occurrence. In fifth and sixth place, shoppers have used the words "website" and "find" respectively. Ful-

ly satisfied customers have used the words “prices”, “quality”, “recommend” and “looking” from seventh to tenth place in terms of frequency. In Figure 3, the top thirty most frequent words are shown.

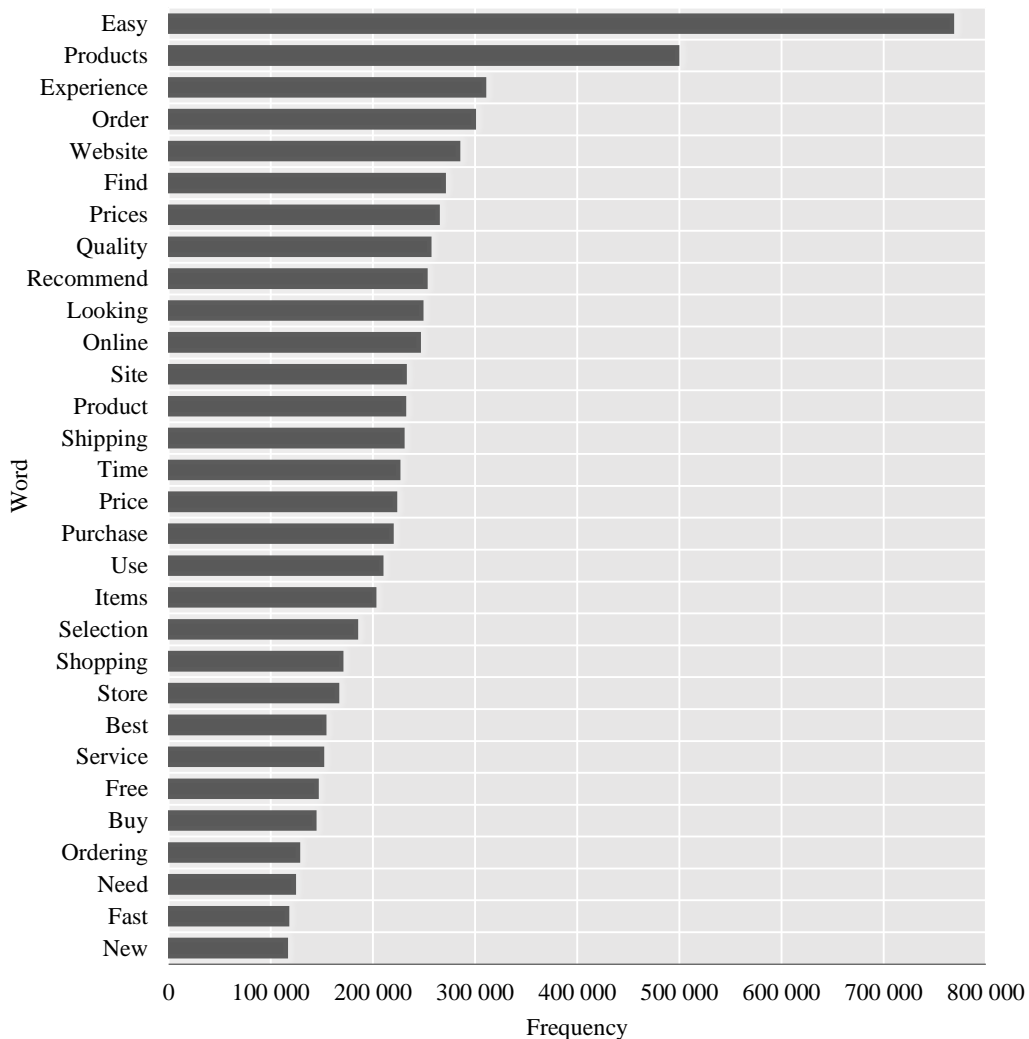


Figure 3. Top thirty most frequent words

In this section, the authors use word cloud visualization techniques to better comprehend the most frequently occurring words. Word clouds are visual representations of word frequency that give more prominence to words that occur more frequently in the source text. As shown in Figure 4, the most frequent word “easy” appears with a larger font size.



Figure 4. World cloud for top thirty most frequent words

Latent dirichlet allocation. In the analysis section, the authors rationalized that topic coherence in this study is the primary approach to determining the number K in topic modeling. Based on the results of the previous section, the authors decided to run LDA modeling for K equals two and three to discover the most appropriate number of topics for customer reviews. Figure 5 (a, b) illustrates the probabilities of words belonging to topics in the first LDA modeling.

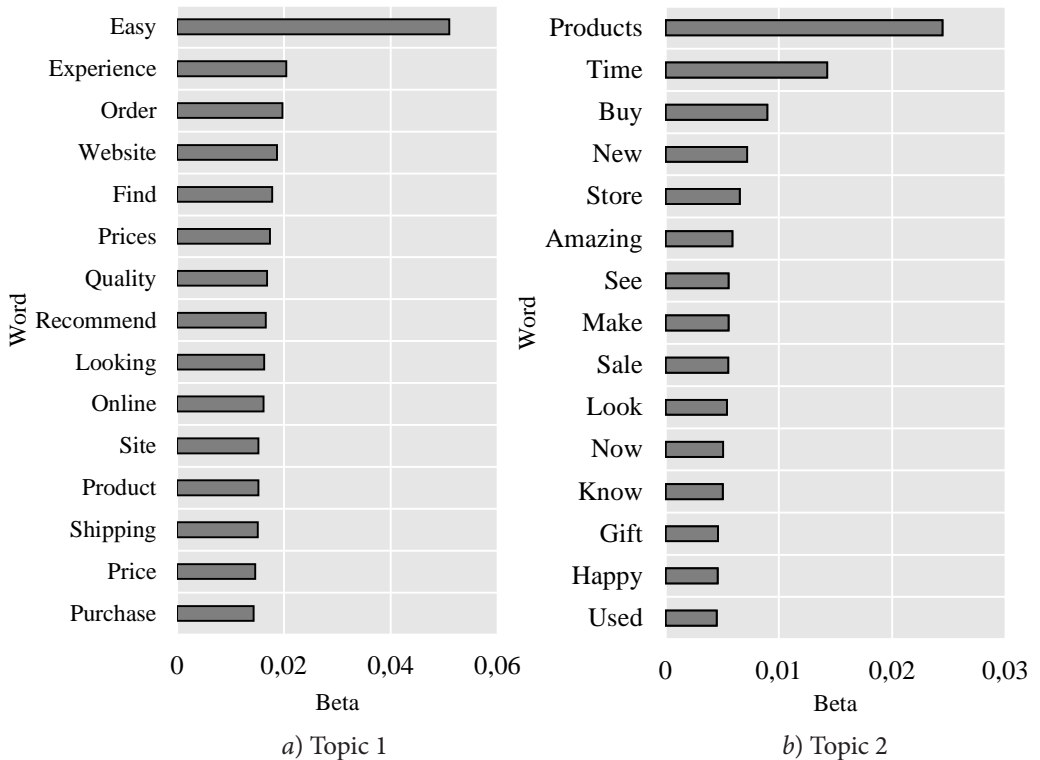


Figure 5. Beta for words in Topic 1 and Topic 2 for LDA model, $K = 2$

For example, the probability that the word “easy” belongs to topic 1 is 5.1%. The other most probable words in this topic are “experience”, “order”, and “website” with a probability of 2.04%, 1.98%, and 1.87%, respectively.

The Topic 2 in the first LDA modeling includes the words “products”, “time”, “buy”, “new”, and “store” as the five most probable words in this topic. For instance, the word “products” occurs as the first most probable word in this topic with a probability of 2.4%. As a result of looking at the most probable words in the first LDA modeling, it can be concluded that the first topic is mostly related to factors such as ease of use, product, and shipping. Nevertheless, the second topic in this LDA modeling mostly exhibits factors related to product, with the first most likely word being “products”. The nature of LDA topic modeling shows that the same word “products” can occur in both topics with different probability values. Thus, topic coherence in both Topic 1 and Topic 2 of LDA modeling with $K = 2$ is low, and the authors opted to try $K = 3$ in order to have better coherence between the terms in each topic.

The second LDA modeling with $K = 3$ reveals the most probable words as shown in Figure 6 (a, b, c).

The first topic includes the words “easy”, “experience”, “website”, “find”, and “prices” as the top most probable words in this topic. The probability of the first word in this topic for the word “easy” is 7.8%. The second topic ties in with the most possible words, such as “products”, “buy”, “new”, “amazing” and “sale”. Finally, the third topic relates mainly to the words “order”, “shipping”, “purchase”, “free”, and “time” as the most likely words in this topic.

In the second LDA topic modeling with $K = 3$, topics differentiate themselves more clearly as compared to the first LDA modeling with $K = 2$. In this modeling, the third topics have a more distinct boundary in showing the most probable words used by highly satisfied customers about their online purchasing experience during the COVID-19 era. Thus, topic coherence in LDA modeling with $K = 3$ is higher compared to $K = 2$, and topic labeling with help of two experts in the field is plausible with the twenty most probable terms in each topic. The first topic in this LDA modeling mainly deals with the ease of use for highly satisfied customers in the COVID-19 era. The words “easy”, “experiment”, “website”, and “find” as the four most likely words to appear in this topic, along with other related words concerning the ease of use. The second topic relates to a factor relating to product characteristics such as quality. In this topic, the word “products” is most often mentioned in the reviews of highly satisfied customers about their online shopping experience in the COVID-19 era. Thus, this topic is labeled as “product”. Finally, the last topic apparently refers to the delivery of purchased products. In this topic, the words “shipping”, “free”, and “time” are mentioned together, which may indicate the prominence of product delivery in terms of customer satisfaction with their online shopping experience. Nevertheless, the third topic is labeled as “delivery”.

Table 3 shows the beta value for the most probable words in each of the three identified topics.

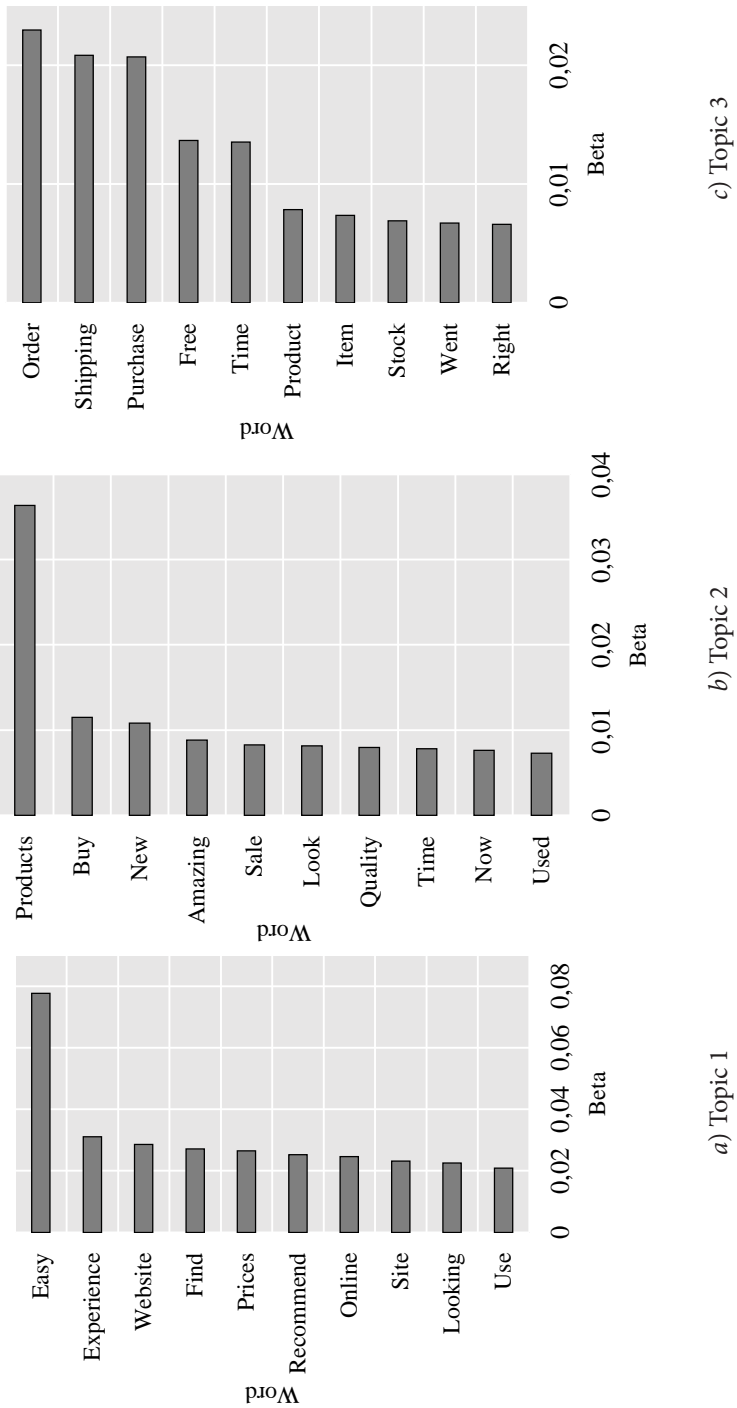


Figure 6. Beta for words in Topic 1, Topic 2 and Topic 3 for LDA model, $K = 3$

Table 3. Beta values for terms for LDA model, $K = 3$

No	Topic	Term	Beta
1	2	3	4
1	1	Easy	0.078
2	2	Products	0.036
3	1	Experience	0.031
4	1	Website	0.029
5	1	Find	0.027
6	1	Prices	0.026
7	1	Recommend	0.025
8	1	Online	0.025
9	1	Site	0.023
10	3	Order	0.023
11	1	Looking	0.023
12	1	Use	0.021
13	3	Shipping	0.021
14	3	Purchase	0.021
15	1	Items	0.020
16	1	Selection	0.018
17	1	Quality	0.017
18	1	Shopping	0.017
19	1	Store	0.016
20	1	Price	0.016

Customer satisfaction factors in online retail: Online review analysis

21	1	Service	0.015
22	1	Product	0.015
23	3	Free	0.014
24	3	Time	0.014
25	1	Ordering	0.013
26	1	Products	0.012
27	1	Fast	0.012
28	2	Buy	0.012
29	1	Customer	0.011
30	2	New	0.011
31	1	Quick	0.010
32	1	Friends	0.010
33	1	Navigate	0.010
34	1	Shop	0.010
35	1	Parts	0.009
36	2	Amazing	0.009
37	1	Nice	0.009
38	2	Sale	0.008
39	2	Look	0.008
40	1	Definitely	0.008
41	1	Check	0.008
42	1	Super	0.008
43	2	Quality	0.008
44	1	Best	0.008
45	3	Product	0.008

End of Table 3

1	2	3	4
46	2	Time	0.008
47	1	Excellent	0.008
48	1	Need	0.008
49	2	Now	0.008
50	1	Checkout	0.008

As mentioned earlier, the authors in this study have done topic labeling on the basis of considering at least the twenty most probable words in each topic based on the probability values of the terms.

DISCUSSION AND CONCLUSION

The findings of this study contribute to the literature on studies involving big text data analysis to thoroughly investigate customer reviews about online purchasing from online retailers in the COVID-19 era. Based on the results of this study, LDA modeling with $K = 3$ has revealed the main topics in customer's reviews in this research. The authors determined the coherence of these topics with the twenty most probable words in each of them as well as with the guidance of two experts in the field. The three main topics concerning the ease of use, product, and delivery have been articulated based on reviews of highly satisfied customers in the COVID-19 era. The authors argue that because of the long period of the COVID-19 era and its ongoing consequences around the world, these three main factors of customer satisfaction, which are identified using topic modeling and the topic coherence approach, may shed light for practitioners and researchers in the context of online retailing, especially during this era.

Al-Ghraibah's findings about the importance of ease of use of a website for customers in online shopping during the COVID-19 era support the outcome of this study [Al-Ghraibah, 2020]. P. Merugu and V. Mohan made similar conclusions related to the importance of the ease of use in the COVID-19 era [Merugu, Mohan, 2020]. In accordance with the second topic related to product, P. Brandtner with co-authors also found that product-related features such as product availability mainly stem from customer sentiment in the COVID-19 era [Brandtner et al., 2021]. The result of this study on the importance of delivery is consistent with a finding by other researchers who discovered that paying attention to consumers' issues throughout product delivery can make them happy and loyal to an online business [Abdallah, Alyafai, Ibrahim, 2021; Kim, Yoo, 2021]. Nevertheless, the results of this article possess a certain originality concerning the introduction of new items defining factors of customer satisfaction in the online shopping. For

instance, the terms “find”, “selection”, “ordering” and “navigate”, which occur in the topic “ease of use”, can help researchers design their questionnaire as well as build a scale measurement based on these items, which may potentially build a latent variable rigorously.

The authors argue that because of the necessity of utilizing large amounts of data in today’s digital world to make timely decisions and particularly the utilization of customers’ online reviews, the findings and methods of this study are very beneficial for online retailing platforms. Online store executives and practitioners can use the outcomes of this research to consistently analyze customer reviews across their platforms using a similar approach and gain a deeper understanding of customers’ perceptions about their products and services. More specifically, online retail executives ought to take a business intelligence approach for relevant departments, such as marketing, that communicates more with customers and build a machine learning algorithm, with the help of an IT department, based on this research’s data analysis method to analyze customers’ reviews thoroughly and gain insights into their online shopping experience. Thus, they can grasp the factors that are very influential to customers in terms of their online shopping satisfaction, as well as address the shortcomings, to make shopping more enjoyable for their customers. Regarding ease of use, online retailers can create optimized web pages along with the main touchpoints of customer interaction and make them more user-friendly. Similarly, in the COVID-19 era, online retailers need to improve the quality of products and delivery. Nonetheless, management implications of this study may be extended to other situations similar to the COVID-19 pandemic, where customers tend to create more user-generated content, such as online reviews.

This study has several limitations. The study is based on the English-language context of US online retailers, and therefore the findings may not be generalizable beyond this context. The authors recommend extending the geographical scope of this study in further research in order to increase the generalizability of the study results. Another limitation of this study is lack of demographic information about customers, such as gender, age, and income which could make the findings of this study more actionable. For further investigation, one may consider analyzing customer feedback considering their gender from the name of the author. For further investigation, the authors intend to adopt more advanced text mining methods to gain insight into not only fully satisfied customers’ reviews but also other segments of customers who submitted reviews. Moreover, increasing the number of K in LDA modeling and clustering customers based on their level of satisfaction derived from their reviews will also be practical for future research.

References

- Abdallah N., Alyafai H., Ibrahim A. 2021. Customer satisfaction towards online shopping. *International Journal of Current Science Research and Review* 4 (7): 692–696.
- Al-Ghreibah O. B. 2020. Online consumer retention in Saudi Arabia during COVID-19: The moderating role of online trust. *Journal of Critical Reviews* 7 (9): 2464–2472.
- Al-Jahwari N. S., Khan F. R., Al Kalbani G. K., Al Khansouri S. 2018. Factors influencing customer satisfaction of online shopping in Oman: Youth perspective. *Humanities & Social Science Reviews, EISSN* 6 (2): 2395–7654.

- Atul Khedkar S., Shinde S. K. 2019. Customer review analytics for business intelligence. In: Krishnan N., Karthikeyan M. (eds). *2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*. Madurai, India: IEEE; 1–5.
- Awad N. F., Ragowsky A. 2008. Establishing trust in electronic commerce through online word of mouth: An examination across genders. *Journal of Management Information Systems* **24** (4): 101–121.
- Bahari N., Samad N. S. A., Yaziz Mohamad M. F. A., Yunoh M. N. M., Rosli N. A. 2021. Factors influencing customer satisfaction in online shopping. *Journal of Entrepreneurship and Business* **9** (1): 72–82.
- Bansal H. S., Voyer P. A. 2000. Word-of-mouth processes within a services purchase decision context. *Journal of Service Research* **3** (2): 166–177.
- Barnes S. J. 2020. Information management research and practice in the post-COVID-19 world. *International Journal of Information Management* **55**: 102175.
- Barnes S. J., Diaz M., Arnaboldi M. 2021. Understanding panic buying during COVID-19: A text analytics approach. *Expert Systems with Applications* **169**: 114360.
- Barnes S. J., Mattsson J., Sørensen F., Jensen J. F. 2020. Measuring employee-tourist encounter experience value: A big data analytics approach. *Expert Systems with Applications* **154**: 113450.
- Berger J., Humphreys A., Ludwig S., Moe W. W., Netzer O., Schweidel D. A. 2020. Uniting the tribes: Using text for marketing insight. *Journal of Marketing* **84** (1): 1–25.
- Blei D. M. 2012. Probabilistic topic models. *Communications of the ACM* **55** (4): 77–84.
- Brandtner P., Darbanian F., Falatouri T., Udokwu C. 2021. Impact of COVID-19 on the customer end of retail supply chains: A big data analysis of consumer satisfaction. *Sustainability* **13** (3): 1464.
- Brosnan K., Babakhani N., Dolnicar S. 2019. “I know what you’re going to ask me”: Why respondents don’t read survey questions. *International Journal of Market Research* **61** (4): 366–379.
- Cao Y., Gruca T. S., Klemz B. R. 2003. Internet pricing, price satisfaction, and customer satisfaction. *International Journal of Electronic Commerce* **8** (2): 31–50.
- Chang H. H., Chen S. W. 2009. Consumer perception of interface quality, security, and loyalty in electronic commerce. *Information & Management* **46** (7): 411–417.
- Deyalage P., Kulathunga D. 2020. Exploring key factors for customer satisfaction in online shopping: A systematic literature review. *Vidyodaya Journal of Management* **6** (1): 163–190.
- Forster P. W., Tang Y. 2005. The role of online shopping and fulfillment in the Hong Kong SARS crisis. In: *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*; 271. IEEE.
- Ghandour A., Woodford B. J. 2020. COVID-19 impact on E-Commerce in UAE. In: *21st International Arab Conference on Information Technology (ACIT)*/ Giza, Egypt: IEEE; 1–8.
- Guo Y., Barnes S. J., Jia Q. 2017. Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management* **59**: 467–483.
- Hashem T. N. 2020. Examining the Influence of COVID 19 Pandemic in changing customers’ orientation towards e-shopping. *Modern Applied Science* **14** (8): 59–76.
- Hernández-Méndez J., Muñoz-Leiva F., Sánchez-Fernández J. 2015. The influence of e-word-of-mouth on travel decision-making: Consumer profiles. *Current Issues in Tourism* **18** (11): 1001–1021.
- Ho-Dac N. N., Carson S. J., Moore W. L. 2013. The effects of positive and negative online customer reviews: Do brand strength and category maturity matter? *Journal of Marketing* **77** (6): 37–53.
- Hsu S.-H. 2008. Developing an index for online customer satisfaction: Adaptation of American Customer Satisfaction Index. *Expert Systems with Applications* **34** (4): 3033–3042.
- Kim J., Jin B., Swinney J. L. 2009. The role of retail quality, e-satisfaction and e-trust in online loyalty development process. *Journal of Retailing and Consumer Services* **16** (4): 239–247.
- Kim R. Y. 2020. The impact of COVID-19 on consumers: Preparing for digital sales. *IEEE Engineering Management Review* **48** (3): 212–218.
- Kim S.-H., Yoo B.-K. 2021. Topics and sentiment analysis based on reviews of omni-channel retailing. *Journal of Distribution Science* **19** (4): 25–35.

- LaValle S., Lesser E., Shockley R., Hopkins M. S., Kruschwitz N. 2011. Big data, analytics and the path from insights to value. *MIT Sloan Management Review* **52** (2): 21–32.
- Liu X., He M., Gao F., Xie P. 2008. An empirical study of online shopping customer satisfaction in China: A holistic perspective. *International Journal of Retail & Distribution Management* **36** (11): 919–940.
- Losiewicz P., Oard D. W., Kostoff R. N. 2000. Textual data mining to support science and technology management. *Journal of Intelligent Information Systems* **15** (2): 99–119.
- Merugu P., Mohan V. K. 2020. Customer satisfaction towards online shopping with reference to Jalandhar city. *International Journal of Management* **11** (2): 36–47.
- Oliveira M., Tavares F., Diogo A., Ratten V., Santos E. 2021. The importance of e-commerce and customer relationships in times of COVID-19 pandemic. In: Ratten V., Thaichon P. (eds). *COVID-19, Technology and Marketing*. Singapore: Palgrave Macmillan; 33–58.
- Oliver R. L. 1977. Effects of expectation and disconfirmation on post-exposure product evaluations: An alternative interpretation. *Journal of Applied Psychology* **62** (4): 480–486.
- Oliver R. L. 1980. A cognitive model of the antecedents and retail setting. *Journal of Retailing* **57**: 25–48.
- Pantano E., Pizzi G., Scarpi D., Dennis C. 2020. Competing during a pandemic? Retailers' ups and downs during the COVID-19 outbreak. *Journal of Business Research* **116**: 209–213.
- Podsakoff P. M., MacKenzie S. B., Lee J.-Y., Podsakoff N. P. 2003. Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology* **88** (5): 879.
- Prasad R. K., Srivastava M. K. 2021. Switching behavior toward online shopping: Coercion or choice during Covid-19 pandemic. *Academy of Marketing Studies Journal* **25** (1): 1–15.
- Reibstein D. J. 2002. What attracts customers to online stores, and what keeps them coming back? *Journal of the Academy of Marketing Science* **30** (4): 465–473.
- Salem M. A., Nor K. M. 2020. The effect of COVID-19 on consumer behaviour in Saudi Arabia: Switching from brick and mortar stores to E-Commerce. *International Journal of Scientific & Technology Research* **9** (07): 15–28.
- Spence M. 1973. Job market signaling. *Quarterly Journal of Economics* **87** (3): 355–374.
- Szymanski D. M., Hise R. T. 2000. E-satisfaction: An initial examination. *Journal of Retailing* **76** (3): 309–322.
- Tandon U., Kiran R. 2019. Factors impacting customer satisfaction: An empirical investigation into online shopping in India. *Journal of Information Technology Case and Application Research* **21** (1): 13–34.
- Vakulenko Y., Shams P., Hellström D., Hjort K. 2019. Online retail experience and customer satisfaction: The mediating role of last mile delivery. *The International Review of Retail, Distribution and Consumer Research* **29** (3): 306–320.
- Westbrook R. A. 1987. Product/consumption-based affective responses and postpurchase processes. *Journal of Marketing Research* **24** (3): 258–270.
- Yang Z., Cai S., Zhou Z., Zhou N. 2005. Development and validation of an instrument to measure user perceived service quality of information presenting web portals. *Information & Management* **42** (4): 575–589.
- Yuen K. F., Wang X., Ma F., Li K. X. 2020. The psychological causes of panic buying following a health crisis. *International Journal of Environmental Research and Public Health* **17** (10): 3513.
- Zhang K. Z. K., Zhao S. J., Cheung C. M. K., Lee M. K. O. 2014. Examining the influence of online reviews on consumers' decision-making: A heuristic-systematic model. *Decision Support Systems* **67**: 78–89.
- Zeng F., Hu Z., Chen R., Yang Z. 2009. Determinants of online service satisfaction and their impacts on behavioural intentions. *Total Quality Management* **20** (9): 953–969.

Received: March 1, 2022

Accepted: December 12, 2022

Contact information

Mehran Haddadi — Postgraduate Student; mhaddadi@hse.ru

Vera A. Rebiazina — PhD in Economics, Associate Professor; rebiazina@hse.ru

**ФАКТОРЫ УДОВЛЕТВОРЕННОСТИ ПОКУПАТЕЛЕЙ В ОНЛАЙН-РЕТЕЙЛЕ:
АНАЛИЗ ОНЛАЙН-ОТЗЫВОВ**

М. Хаддади, В. А. Ребязина

Национальный исследовательский университет «Высшая школа экономики»,
Российская Федерация, 101000, Москва, ул. Мясницкая, 20

Для цитирования: Haddadi M., Rebiazina V. A. 2023. Customer satisfaction factors in online retail: Online review analysis. *Вестник Санкт-Петербургского университета. Менеджмент* 22 (1): 3–22. <http://doi.org/10.21638/11701/spbu08.2023.101>

В статье представлены результаты исследования факторов удовлетворенности клиентов, основанного на отзывах потребителей об интернет-магазинах США. Созданный потребителями контент приобрел особую информативность в эпоху COVID-19 в результате быстрого роста числа покупок, совершенных онлайн. Используемая в исследовании база данных состоит из 5 340 786 онлайн-отзывов и получена с помощью агрегатора потребительских отзывов bizratesurvey.com. В работе рассматриваются индивидуальные отзывы клиентов о 839 интернет-магазинах США. Метод частотного анализа слов и скрытое распределение Дирихле использованы как основные методы анализа полученных данных. В результате было выявлено три основных блока — «Простота использования», «Продукт» и «Доставка», — которые упоминались высокоудовлетворенными клиентами в отзывах об интернет-магазинах. На основе выводов исследования руководители онлайн-магазинов могут принимать решения об улучшении сервиса и доставки за счет простоты потребления и повышения качества продукции. Кроме того, специалисты-практики могут воспроизвести разработанную методологию для анализа отзывов клиентов и экстраполировать ее на другие исследования. Полученные результаты также отражают новый способ использования больших текстовых данных для аналитики отзывов клиентов в академических исследованиях.

Ключевые слова: COVID-19, потребительская удовлетворенность, онлайн-ритейл, текстовый анализ, онлайн-отзывы.

Статья поступила в редакцию 1 марта 2022 г.

Статья рекомендована к печати 12 декабря 2022 г.

Контактная информация

Хаддади Мехран — аспирант; mhaddadi@hse.ru

Ребязина Вера Александровна — канд. экон. наук, доц.; rebiazina@hse.ru

Исследование выполнено в ходе реализации фундаментального исследовательского проекта «Цифровизация как движущая сила открытых инноваций и совместного творчества: последствия для создания ценности» в рамках Программы Высшей школы бизнес-исследований НИУ ВШЭ 2021–2023 гг. (Протокол № 23 от 22 июня 2021 г. Научной комиссии ВШБ НИУ ВШЭ).