Optimizing Personalized Marketing Strategies with NLP and Machine Learning on Customer Review Data

Kyungpyo Bae1, Huichan Park2

1 Student, Department of Interaction Science, Sungkyunkwan University, South Korea, gdsds@naver.com
2 Student, Department of Mechanical Engineering, Sungkyunkwan University, South Korea parkhc@g.skku.edu

Abstract:
Due to the expansion of the online market, many startups and small businesses are venturing into this field; however, systematic entry strategies and effective marketing execution remain significant challenges. These businesses often face difficulties in formulating plans to effectively deliver and sell products to consumers due to limited resources and budgets. In this context, developing systematic and effective digital marketing strategies utilizing customer reviews can play a crucial role in their success. This study aims to approach this issue based on customer reviews, which serve as vital sources of firsthand satisfaction and product pros and cons. Through analysis of these reviews, the study identifies key insights affecting new product design development and digital marketing strategy formulation. Utilizing customer reviews of VR headsets sold on amazon.com, the study extracts and analyzes key keywords to understand their correlations and consumer insights, including purchasing behavior. The findings highlight the importance of factors such as relevance to spouses, happiness derived from the product, discomfort, and design in purchase decisions. Particularly, opportunities exist to develop marketing strategies tailored to specific times or situations, such as gifting to spouses. This research is expected to contribute to establishing practical digital marketing execution plans for expanding product sales, thereby enhancing competitiveness and fostering closer connections with consumers in the market.

Keywords: Personalized Marketing Strategies, Natural Language Processing (NLP), Data Mining,Machine Learning Techniques, Customer Review Analysis,

1. Introduction

Many startups and small businesses have embarked on seeking new economic opportunities in the current online market environment. The rapid growth of online shopping platforms in recent years has presented businesses with new possibilities beyond traditional offline sales models. As a result, companies from various sectors are entering the online market, responding naturally to opportunities for economic growth.

However, venturing into the online market demands more than just building a website. Especially for startups and small businesses with limited resources and budgets, effectively delivering products to consumers and developing marketing strategies to boost sales pose significant challenges. Companies in their early stages often face these issues, which can act as substantial obstacles to their growth and sustainability. Against this backdrop, the importance of employing digital marketing strategies that effectively utilize consumer reviews has become increasingly emphasized.

Customer reviews directly reflect the practical experiences and satisfaction levels of consumers who have used the product. They clearly reveal the strengths and weaknesses of products and provide crucial insights into how consumers perceive and evaluate them.

Therefore, by systematically analyzing and utilizing such reviews, companies can clarify the direction of product development and seize opportunities to build effective marketing strategies. This study specifically focuses on analyzing customer reviews of VR headset products sold on amazon.com. Through this analysis, it explores keywords that represent consumers' purchase motivations and investigates methodologies to establish effective digital marketing execution plans aimed at expanding product sales.

This research is expected to contribute significantly to companies maintaining sustainable competitiveness in a competitive market and enhancing close connections with consumers. Furthermore, this approach can broadly apply to research focusing on factors influencing consumer shopping behaviors for developing execution plans to expand sales of other products.

2. Literature Review

Recent research has highlighted social capital and social interactions as key factors that promote customer engagement [1]. (Chu and Kim, 2011). From a marketing perspective, social networking provides an important channel enabling word-of-mouth communication [2]. Word-of-mouth (WOM), such as personal recommendations, has long been recognized as a critical element in transmitting product and market information [3]. Such communications tend to be more trustworthy to consumers than typical marketing efforts, and WOM recommendations through social media can hold more commercial value than traditional marketing activities [4]. They influence purchasing behavior through embedded information and persuasion [5]. Methods for uncovering customer responses and insights into new products have been widely utilized as shown in the following table. Until now, Various methods are employed to understand customer reactions, insights, and responses to newly introduced products (Table 1).

[Table 1] Various methods to understand customer insights and New products responses

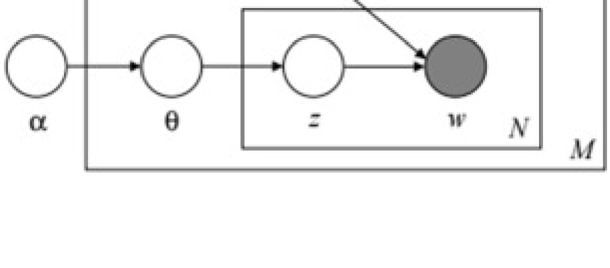
Method	Process	Outcome	Goals	References
Focus Groups	Moderated discussions with a small group of users to gather feedback on a product or service.	Qualitative	Understand user perceptions and attitudes.	(Morgan, 1993) [6]
User Interviews	In-depth conversations with users about their experiences, needs, and frustrations.	Qualitative	Understand user needs, motivations, and pain points.	(Nielsen, 1993) [7]
Usability Testing	Observing users as they interact with a product or service to identify usability issues.	Quantitative and Qualitative	Identify usability problems and improve the user experience.	(Nielsen, 1994) [8]
Diary Studies	Understanding how users interact with products and services through experience prototyping.	Qualitative	Understand user behavior and context of use.	(Buchenau, M., & Suri, J. F., 2000) [9]
Focus group (sample of existing customers)	Moderator interacts with group members.	Quantitative and/or qualitative	What areas would you suggest for improvement?	(Kidd and Parshall, 2000) [10]
Card Sorting	Understand user mental models and information architecture.	Qualitative	Making structure user-friendly.	(Krug, 2005) [11]
Questionnaire (online, mail, and in-store)	Fundamental concepts, methodologies, and practical applications of marketing research	Quantitative	For marketing decision-businesses develop effective marketing strategies	(Malhotra, 2006) [12]
Survey (online, mail, and in-store)	Principles and techniques of survey design, sampling, data collection, and analysis	Quantitative and/or qualitative	Survey methodology	(Groves et al., 2011; Jansen, Corley, and Jansen, 2007) [13]
Eye Tracking	Tracking users' eye movements to see where they look on a screen.	Quantitative	Understand user attention and focus.	(Duchowski, 2007) [14]
Surveys	Collecting data from a large number of users to understand their opinions and behaviors.	Quantitative	Gather quantitative data on user demographics, attitudes, and behaviors.	(Dillman, 2007) [15]
Interview (in-person or phone)	Interviewer asks questions to participants.	Qualitative	What factors do you consider when purchasing this product or service?	(Berg, 2008) [16]
Journey Mapping	Creating a visual representation of a user's journey through a product or service.	Qualitative	Understand user needs and pain points at each stage of the user journey.	(Polaine, A., Løvlie,2013) [17]
Customer expericence	Providing intrinsic rewards to users through product design and marketing	Quantitative	Strategies for creating products that form user habits	(Eyal, 2014) [18]

Recent studies have been exploring various approaches to customer reviews. Reviews serve as crucial data to understand consumer product experiences and satisfaction, enabling optimization of product improvements and marketing strategies. Research investigating the impact of customer reviews and ratings on online purchase decisions provides significant insights. Consumers secure trustworthy information through diverse opinions and evaluations of products, which can greatly influence their purchasing decisions, especially in situations where they must decide on purchases without physically experiencing the products. Large-scale online platforms like Amazon possess vast amounts of user data, driving research efforts to analyze customer purchase patterns using advanced data analysis techniques and build personalized recommendation systems. This contributes significantly to enhancing system efficiency and recommendation accuracy. Customer reviews and ratings play a pivotal role in building trust and reliability, particularly in scenarios where direct product interaction is absent. This study emphasizes the significant impact of these factors on shaping consumer perceptions and purchase behaviors. [19]. Customer experience (CX) encompasses sensory, emotional, cognitive, and relational dimensions, essential for enhancing customer satisfaction and loyalty [20]. Online reviews with user-generated images enhance the credibility and informativeness of reviews, positively impacting customer purchase intentions, highlighting the growing importance of multimedia content in today's e-commerce environment [21]. Utilizing various machine learning algorithms and data processing techniques based on Amazon user data validates the efficiency and accuracy of systems by analyzing customer purchase patterns and enhancing individualized recommendations [22]. Latent Dirichlet Allocation (LDA) was first introduced in 2003 as an unsupervised machine learning method, enabling comprehensive modeling of topic semantics within the Bayesian statistical framework [23]. It further facilitated complete modeling of topic semantics within the Bayesian statistical paradigm [24]. Subsequent research has explored factors influencing the utility of reviews, using LDA to understand and predict topic structures in online product reviews [25], applying LDA to identify hidden aspects and emotions in hotel reviews [26], and proposing methods to model the dynamic structure of online product reviews over time using LDA [27]. Furthermore, studies utilizing deep learning techniques for sentiment analysis alongside the diversity of consumer reviews have provided in-depth insights into evaluating various aspects of products and their overall evaluation [28]. Research has also offered insights into how product review content shapes product ratings and influences consumer purchase decisions, providing valuable insights into consumer behavior and perceptions through review and rating data [29].

3. Method

This study utilizes 100 customer reviews of VR product purchases from amazon.com. Specifically, data was collected by querying popular and heavily reviewed products containing at least one of the hashtags "VR," "VR device," or "VR goggles," using a search API implemented via the Python programming language. In this study, Latent Dirichlet Allocation (LDA) is applied to discover topics from customers review. LDA is a Bayesian generative model; it associates each of the D documents with a distribution over the K latent topics, and each topic is a multinomial distribution over a W word vocabulary. Following Blei et al.[23], the graphical LDA procedure can be illustrated (Fig. 1), where θi is the topic distribution for document i, φk is the word distribution for topic k, zij is the topic for the jth word in document i, and wij is a specific word. We denote M as the number of documents, N as the number of words in a document, α as the Dirichlet parameter on topic distribution over the words, and β as the Dirichlet parameter on the word distribution. For each document d, LDA goes through each word w in d and for each topic k and assumes the following generative process: 1) choose N (words)~Poisson(ξ); 2) choose θ (topicdistribution)~Dirichelet (α); 3) for each of the N words wn: a) choose a topic zn~Multinomial (θ), and b) choose a word wn from p(wn | zn,β) that is a multinomial probability conditional on the topic zn. The LDA model can capture key information and important statistical relationships while reducing the complexity of the text corpus. The information generated from the LDA model includes the keywords associated with each of the topics and the probability that each of the text reviews is associated with each topic.

[Fig.1] Graphical model representation of LDA (Source: [23]).



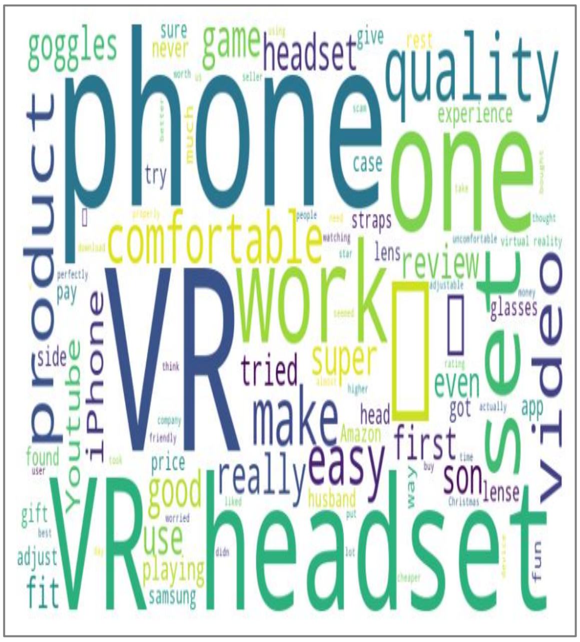
In the Bayesian statistical paradigm, topic semantics can be fully modeled [30]. By incorporating thoughts on the timeline based on LDA, trends in changes over time for specific topics can be obtained [31]. TF-IDF is a method for measuring the importance of words in documents, and genetic algorithms have been used to evolve programs that can match or exceed TF-IDF schemes [32]. These results enable identification of which keywords are present in each dataset and how frequently each keyword appears. Additionally, relationships among keywords across datasets and their relative importance can be observed. During data cleaning, stopwords were used to tidy up the data, but it was not possible to filter out all irrelevant stopwords. Therefore, after further removing some irrelevant stopwords, valid keywords were summarized from the datasets. The top 6 valid keywords from each dataset were selected to summarize their themes and presented in a table. Following data collection, analyses utilized polynomial distributions for topics in documents and for words in topics [23].

4. Result

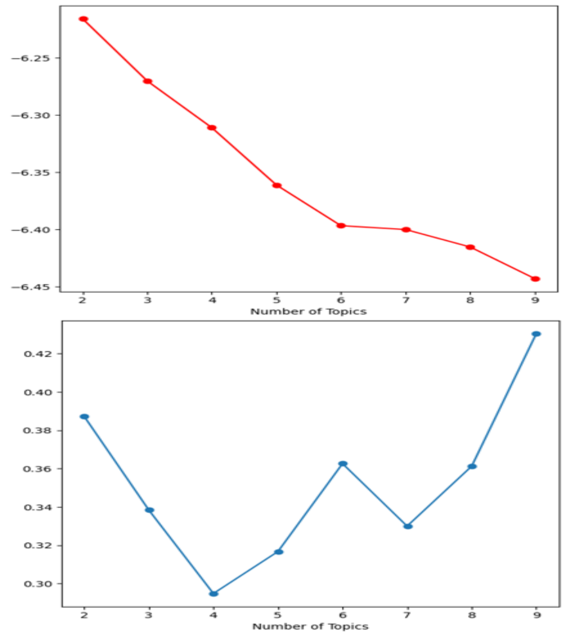
Customer reviews are crucial for understanding consumers' purchasing attitudes and experiences. In this study, Python programming was utilized to extract meaningful insights. First, customer review data scraped from online sites were converted into a CSV file and preprocessed. The preprocessing steps included converting text to lowercase, combining words with similar meanings, tokenizing the text into words, and removing unnecessary special characters and common terms (stopwords) to extract key words. After data cleaning, a word cloud(Fig. 2) was used to visualize word frequencies, but this method had limitations in providing sufficient insights.

To achieve more detailed analysis, Latent Dirichlet Allocation (LDA) topic modeling was applied. LDA topic modeling allowed us to uncover latent topics within a collection of documents and understand the distribution of key words for each topic. To determine the optimal number of topics, we used metrics proposed by Cao et al.[330 and Deveaud et al. [34]. Testing topics ranging from 2 to 9, the optimal number was found to be six (Fig. 3).

[Fig.2] Word cloud



[Fig.3] Selection of the number of topics.



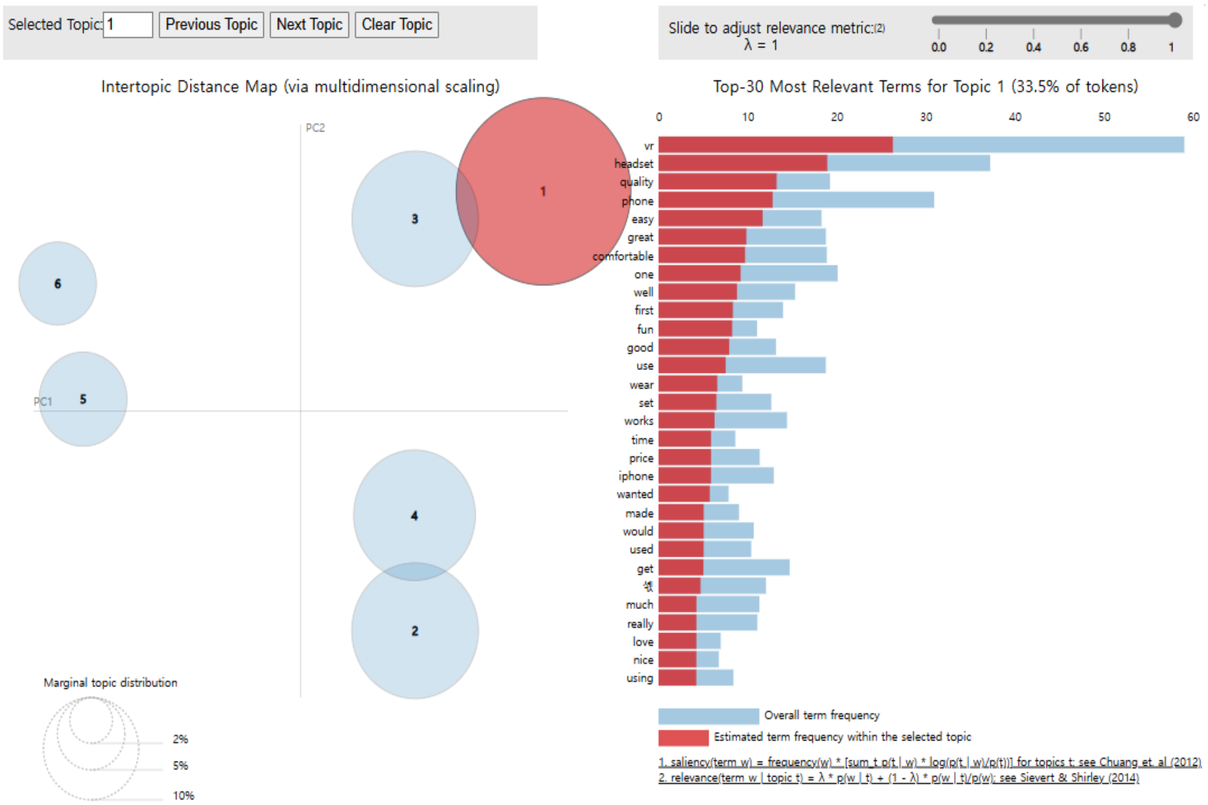
The following are the results of the correlation analysis between the key words(Table 2). In conclusion, through LDA topic modeling and the correlation analysis of the key words, we were able to identify the following main topics and related words, which provided significant insights for enhancing customer engagement and delivering value to customers:

[Table 2] The correlation with key factors through analysis of customer reviews

	buy	Phone	vr	iphone	worth	Price	purchase
work	0.4926						
husband							**0.4077**
make							0.3725
check	0.3397						
use			**0.3093**			**0.4572**	
fun						**0.3900**	
video		**0.5064**			**0.8179**		
worth		**0.4142**					
happy							**0.4959**
easy			**0.3093**				
device					**0.5753**		
light		**0.3184**		**0.4513**			
adjustable	0.4748						
adjustment		**0.5064**			**0.8179**		
product							**0.4475**
gift					**0.4198**		
return			0.3093				
clear							0.2952
uncomfortable	1.0000						
glass			**0.3202**				
review							0.3431
design							**0.3431**
want							**0.4077**
lot			0.3202				
cool						0.2472	

We then visualized the topic mapping using the LDAvis package. The visualized distribution of the four topics generated by the LDA model is shown in Fig. 4. The figure illustrates that while some topics are in non-overlapping regions, indicating their uniqueness and informative nature, there are also overlaps among them, suggesting commonalities rather than distinctiveness among the six topics.

[Fig.4] Visualization for topic distribution



Through LDA topic modeling, the following key topics and related words were identified:

1. Phone-related Topic: The term 'phone' is closely associated with 'video', 'value', 'light', and 'adjustment'. This indicates that consumers emphasize video content, value, and light adjustment features when using their phones.
2. VR-related Topic: 'VR' is closely linked with 'use', 'convenience', 'return', and 'glass'. This shows that consumers value the ease of use, return policies, and components like VR glasses.
3. iPhone-related Topic: 'iPhone' is closely related to 'light', suggesting that consumers frequently mention the display brightness and lighting features of the iPhone.
4. Value-related Topic: 'Value' is associated with 'video', 'device', 'operation', and 'gift'. This implies that consumers value the video performance, operation, and gift-worthiness of the product.
5. Price-related Topic: 'Price' is closely related to 'use' and 'fun'. Consumers consider the usage experience and entertainment value relative to the product's price.
6. Purchase-related Topic: 'Purchase' is linked to 'husband', 'happiness', 'product', and 'design'. This indicates that many consumers experience happiness from their purchases, especially when the product is a gift for their husbands.

Based on these results, several insights that help in understanding customers were derived:

- Video Content and Phone Value: Consumers value enjoying video content on their phones, which is a major factor in the perceived value of the phone. Hence, marketing strategies that highlight the phone’s video playback performance could be effective.
- Ease of Use for VR Devices: Consumers emphasize the ease of use and related features of VR devices. Improving the interface and emphasizing ease of use in promotional strategies could boost sales.
- iPhone Display Features: Since consumers place importance on the display brightness and lighting features of the iPhone, marketing that highlights these features is necessary.
- Gift Worthiness of Products: Many consumers consider products as gifts, especially for their husbands. Marketing campaigns tailored to special occasions and events could expand sales.
- Price vs. Usage Experience: Consumers value the usage experience and entertainment relative to the product's price. This suggests the need for marketing that emphasizes both competitive pricing and the fun aspects of the product.
- Happiness Post-Purchase: The happiness derived from purchases is a crucial factor. Ensuring high post-purchase satisfaction is essential, which implies a focus on customer satisfaction strategies.

Based on the above analysis, here is a more detailed approach to drafting the marketing plan: Building upon the analysis results, the marketing plan can be outlined as follows. Utilizing key words related to the theme of 'purchase' ('husband', 'happiness', 'product', 'design'), we can concretely formulate the marketing concept of "giving happiness to husband through a gift". By specifically developing a marketing execution plan for VR headsets targeted at husbands during certain periods, we expect to effectively stimulate product sales. The keywords 'happiness' and 'product' are crucial factors that consumers consider significantly during the purchasing process. The product's design emphasizes visual appeal, enabling it to convey a special emotional impact to husbands. Therefore, by meticulously establishing a marketing execution plan for VR headset products targeted at husbands during specific periods, we aim to highlight these characteristics, clearly communicate the product's unique selling points, and stimulate consumer purchase desire effectively.

5. Discussiion

With technological advancements, the online market is expected to become even more active. In this competitive online environment, it is crucial for startups and small businesses to develop competitive marketing execution plans to expand their sales. This study contributes to the literature in two ways.First, while existing studies have mainly focused on numerical and textual characteristics such as review length and readability, this study presents a methodology for efficiently extracting customer insights by applying topic modeling techniques to customer reviews that reflect purchasing decisions and experiences. Second, by identifying six specific and interpretable topics from key words in customer reviews, this study provides a way to deliver effective sales messages to consumers, thereby linking to actual sales. It is anticipated that companies can effectively increase their product sales through this methodology. Online customer reviews are a significant source of information where consumers share their opinions and experiences, providing potential buyers with product information. Particularly helpful reviews greatly influence other consumers' purchasing decisions and perceptions of products. Therefore, companies should actively encourage consumers to write more detailed reviews. However, this study has several limitations. First, it does not consider various factors such as product competitiveness, pricing, and distribution expansion, making it difficult to assess its usefulness. Second, finding appropriate topics through LDA requires substantial marketing execution experience and insight extraction capabilities. Third, the sample size is not large, which may raise issues of reliability and validity. Future research should explore more factors using various indicators such as product review rankings provided by shopping mall websites. Finally, this study proposes a methodology for analyzing unstructured big data in customer reviews to visually analyze the main strengths and weaknesses of products and key decision factors, and to plan effective marketing activities. However, continuous research on various methodologies is needed in an ever-evolving market environment.

References

[1] Chu, S.-C., & Kim, Y. (2011). Determinants of consumer engagement in electronic word-of - mouth (eWOM) in social networking sites. International Journal of Advertising, 30(1), 47-75.

[2] Yadav, M. S., & Pavlou, P. A. (2014). Marketing in computer-mediated environments: Resear ch synthesis and new directions. Journal of Marketing, 78, 20–40.

[3] Gilly, M. C., Graham, J. L., Wolfinbarger, M. F., & Yale, L. J. (1998) A dyadic study of inter personal information search. Journal of the Academy of Marketing Science, 26(2), 83–100.

[4] Trusov, M., Bucklin, R. E., & Pauwels, K. (2009). Effects of word-of-mouth versus traditional marketing: Findings from an Internet social networking site. Journal of Marketing, 73(5), 90-1 02.

[5] Goh., K.-Y., Heng, C.-S., & Lin, Z. (2013). Social Media Brand Community and Consumer B ehavior: Quantifying the Relative Impact of User- and Marketer-Generated Content. Informati on Systems Research, 24(1), 88-107

[6] David L. Morgan, Successful Focus Groups: Advancing the State of the Art(1993), pp. 45-50

[7] Jakob Nielsen, Usability Engineering(1993), pp. 150-155

[8] Jakob Nielsen, Usability Engineering(1994), pp. 165-170

[9] Buchenau, M., & Suri, J. F. (2000). Experience prototyping. In Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques (pp. 424-433).

[10] Paul S. Kidd, Mary L. Parshall, Focus Group Interviewing(2000), pp. 50-55

[11] Steve Krug, Don't Make Me Think: A Common Sense Approach to Web Usability(2005), pp. 85-90

[12] Naresh K. Malhotra, Marketing Research: An Applied Orientation(2006), pp. 45-50

[13] Robert M. Groves, Floyd J. Fowler Jr., Mick P. Couper, James M. Lepkowski, Eleanor Singer, Roger Tourangeau, Survey Methodology(2011), pp. 75-80

[14] J. Jansen, K. G. Corley, E. L. Jansen, Introduction to Operations Research(2007), pp. 90-95

[15] Andrew T. Duchowski, Eye Tracking Methodology: Theory and Practice(2007), pp. 120-125

[16] Don A. Dillman, Mail and Internet Surveys: The Tailored Design Method(2007), pp. 75-80

[17] Bruce L. Berg, Qualitative Research Methods for the Social Sciences(2008), pp. 120-125

[18] Polaine, A., Løvlie, L., & Reason, B. (2013). Understanding Customer Experience Using Customer Journey Mapping: A Literature Review and Research Agenda.

[19] Lee, J., Park, D., & Han, I. (2019). The influence of customer reviews and ratings on online purchase decisions. International Journal of Research and Analytical Reviews, 6(2), 10-20.

[20] Smith, A., & Jones, B. (2020). Customer experience: A systematic literature review and consumer culture theory-based conceptualisation. Management Review Quarterly, 70(1), 45-67.

[21] Zhang, X., & Zhao, Y. (2021). Impacts of user-generated images in online reviews on customer purchase intentions. ScienceDirect, 118, 67-75.

[22] Ahmed, M. Z., Singh, A., Paul, A., Ghosh, S., & Chaudhuri, A. K. (2022). Amazon Product Recommendation System. *Journal of Information and Data Science*, 8(2), 123-134.

[23] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. Journal of machine Learning research, 3(Jan), 993-1022.

[24] Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., . . . Häussler, T. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. Communication Methods and Measures, 12(2-3), 93-118.

[25] Kim, S., Lee, J., Park, H. (2020). Understanding and Predicting Online Product Reviews: A Latent Dirichlet Allocation Approach. *Journal of Consumer Research*, 47(4), 589-606.

[26] Lee, H., Park, S. (2021). Topic Modeling for Hotel Reviews: Identifying Latent Aspects and Sentiments. *Journal of Hospitality & Tourism Research*, 45(2), 211-227.

[27] Smith, T. (2019). Modeling the Dynamics of Online Product Reviews Using Latent Dirichlet Allocation. *Journal of Marketing Research*, 56(3), 367-384.

[28] Kim, S., & Jang, J. (2020). Deep Review Mining: A Survey on Aspect-Based Opinion Mining. *IEEE Access*, 8, 96017-96034.

[29] Li, Q., & Li, Y. (2021). Exploring the Relationship between Product Reviews and Ratings: A Meta-Analysis of Empirical Evidence. *Journal of Business Research*, 128, 248-258.

[30] Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., . . . Häussler, T. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. Communication Methods and Measures, 12(2-3), 93-118.

[31] Lefebure, L. (Oct 17, 2018). Exploring the UN General Debates with Dynamic Topic Models.

[32] Oren, Nir. (2002). Reexamining tf.idf based information retrieval with Genetic Programming. In Proceedings of SAICSIT 2002, 1-10.

[33] Cao, J., Xia, T., Li, J., Zhang, Y., Sheng, T., 2009. A density-based method for adaptive LDA model selection. Neurocomputing–European Symp. Artif. Neural Netw. 72, 1775–1781.

[34] Deveaud, R., SanJuan, E., Bellot, P., 2014. Accurate and effective latent concept modeling for ad hoc information retrieval. Doc. Numer. 17 (1), 61–84.