Design of the Multi-purpose Marketing Data Analysis Program: Using actual online sales data

Donggyu Choi[1], Tae-su Wang[2], Minyoung Kim[3], Jongwook Jang[4]

[1]Electronics and Telecommunications Research Institute, Daejeon, Republic of Korea; E-mail: dgchoi@etri.re.kr

[2,3,4]Dong-eui University, Busan, Republic of Korea; E-mail: jwjang@deu.ac.kr

Abstracts: Marketing is a strategy for the sale of goods. Most shopping malls and vendors value this, and it is also a measure of the company's success. To this end, strategies are established through marketing departments or specialized companies, and related data is collected directly. Data analysis uses some tools, but in the end, data must be collected directly, and most of the tools that collect automatically only use SNS networks. In this paper, we present a design method for programs that can be fundamentally utilized without customizing or providing the functions of existing data analysis tools. In the program, HTML structures for commonly used shopping malls, not SNS, are predefined and prepared for immediate use of crawling techniques, and some textual data are generated as Excel files. This can be applied in real time, and some artificial intelligence and emotional analysis are applied to the image to help make an objective evaluation.

Keywords: Marketing, Data Analysis, Tools Program, Artificial Intelligence, Sentiment Analysis.

1. INTRODUCTION

Until now, the field of marketing is often analyzed using human emotions. This is because the things sold are things that people use and are traded according to need. Apart from general calculations, it is an area that requires human decision, but the work required to reach a conclusion is quite time-consuming. Benchmarking is used as a frequently used marketing management strategy. It aims to identify representative strategies of competitors or similar industries, reflect advantages, and exclude disadvantages. However, this is also quickly and easily accessible only when basic data is supported and provided with the time and criteria to analyze it.

Fig. 1. Mask images released by various customer needs.

Recently, a representative example of benchmarking products is the sale of disposable masks due to the coronavirus, and Figure 1 is an image of various masks sold by customers' needs. Existing disposable masks were not sold in the form of competitive markets. In general, the function or performance of the drug is guaranteed according to the manufacturing standards of the drug that meet the national standards, and it was purchased or used where necessary according to the grade. However, based on the serious infectivity of the coronavirus, various companies have manufactured and sold masks [1-4]. Initially, even if a basic type of mask was produced, the supply according to demand could not be resolved, but the demand began to gradually decrease as the mask that had already been purchased did not consume. However, as the situation gradually improved, companies that manufactured various shapes and colors emerged at a time when sales of general masks decreased, and the related market grew exponentially. In addition, unlike general disposable masks, colorful masks are advanced or added functions by entering some areas of fashion, and there are currently a wide variety of products even if they are masks with the same quarantine grade. As such, some products change and improve according to the situation or consumer needs. Even in the area of marketing or management, it is necessary to check and quickly reflect these market trends.

As described above, it is important to read and analyze the market quickly for management strategies, and the process before analysis takes a considerable amount of time. The current analysis process for marketing requires people to collect or organize data themselves, and there are only programs that can digitize it, so there are many wasteful tasks such as directly entering data. Existing marketing data collection tools are generally limited to SNS websites, and there is an inconvenience of having to integrate and check other tools because SNS companies provide them using their own data. Recently, the need for a system such as a benchmark based on the company's own database has been mentioned by regularly monitoring data on consumer goods to be marketed prior to using marketing analysis tools [5].

In this paper, we describe the necessity and design method of basic programs that help facilitate the entire process of performing data analysis so that management strategies such as marketing can be used for online commodity distribution. The structure of representative online shopping sites, not data provided by a single company, can be identified in advance, designed, and used regardless of type, and data from various companies can be collected. The collected data can be used as a graph in Excel, and unlike the commonly used data of research institutes, it is raw, and detailed confirmation is possible. In addition, natural language processing can be performed, and the marketing value of the product can be confirmed through features such as the color of the detailed description image and the atmosphere of reviews by real users of the product. In order for companies to establish strategies similar to benchmarking, costs and time factors such as establishing related departments or performing them through marketing companies can be reduced, and desired data can be directly checked, stored as digitized data, or utilized.

2. RELATED WORK

The size of the online shopping market is huge and there are various types of goods accordingly. The shopping mall and its affiliated vendors compare each other's sales strategies and imitate some of the good strategies. Such indicators refer to product detailed description images or reviews and detect and use the atmosphere accordingly. The program to be designed accordingly was conducted based on the following data.

2.1 E-commerce Market Growth

Online shopping is widely known in other words as e-commerce. E-commerce started in 1982 with the Boston Computer Exchange. It was the world's first e-commerce company, and since then, with the rapid development of the Internet in the 2000s, sales methods and goods have increased dramatically through the Internet, and now it is an era where online purchases are preferred rather than going to a mart [6].

In particular, as it became difficult to go out after the 2020 COVID-19 pandemic, the online market, which remained stagnant to some extent, has grown even more. Overall consumption seemed to decline slightly after the outbreak of COVID-19, but this soon recovered from online consumption, and it was confirmed that more than $1 in

$5 of retail sales was spent online [7,8].

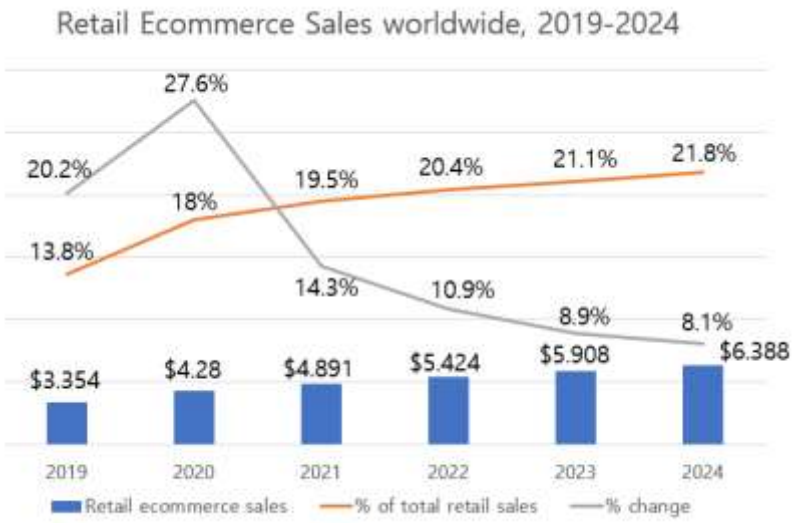

Fig. 2. 2019-2024 Retail E-commerce Sales worldwide.

Figure 2 shows the market size surveyed by eMarketer in relation to the global e-commerce market in 2021. The global e-commerce market in 2021 is estimated to be $13 trillion and 6474 trillion in 2022 [9,10]. According to research firm Euromonitor, the e-commerce market has grown steadily every year, and total sales in 2021 were $788 billion, up 19.3% year-on-year [11]. Through the results of other research institutes, e-commerce is expected to grow steadily, and it is changing more rapidly after the 2020 COVID-19 pandemic.[12]

2.2 Benchmarking

Companies are benchmarking other companies for management strategies that dominate competition. Benchmarking refers to the act of learning according to advantages through comparative analysis of similar cases. On the corporate side, the goal is to compare and analyze the characteristics of other companies' organizations or related products and services to see and learn the advantages. This is usually prominent in terms of marketing, which is often used in market competition. One paper state that benchmarking can be a key learning mechanism that identifies, builds, and enhances marketing capabilities for its success, as well as being performed for sustainable advantage and preemption [13]. Companies should continue to compete with companies that do similar businesses unless they are monopolized, and business efficiency improves when they resolve it through benchmarking [14]. This benchmarking suggests that collecting data is important, and that the ability to effectively collect and analyze a large number of data is needed in all processes related to production as well as sales [15].

2.3 Data crawling and marketing data collection

Data used by companies for marketing is usually collected directly from marketing-related departments or companies. Usually, in the case of a large company, it is possible to perform it through a company because it actually has a marketing department or has sufficient capital. It is difficult to perform professionally for small businesses and startups because it requires a lot of time or money for marketing. In order to solve some of these parts, data analysis is judged and solved by humans, but if data is easily collected, the process is partially simplified. As the concept of big data began to be known, crawling techniques were used, recognizing that it was difficult for people of large-scale data to collect directly. Similar to crawling, it is called scraping and mining, similar but somewhat different in terms of technology. In fact, it was suggested that crawling and scraping in marketing were required, and the collected data could be used for analysis in any form [16]. This data collection method is proving to be convenient and useful through practical success stories of e-commerce operators as well as actual

online marketers [17,18]. However, it is illegal to comply with the policy in the 'robots.txt' file of the relevant Internet site, and to use it for commercial purposes such as receiving and selling money in addition to data analysis and research, or to penalize the original artist.

2.4 Consumption trends according to product image

Various types of data provided by companies exist in different forms. Text in the form of a product name can also be data, but it is most effective to convey what a company wants to introduce with image data. Image is very important in e-commerce, an online market where it is impossible to inspect products directly. Visual imagery is the most powerful way to communicate important information easily and influence choice [19,20]. To create such an image, an entity must consider and work on the interaction of what it wants to market with the color of the logo regarding the company's brand [21]. In fact, it actually reflects what it wants to market through the color that symbolizes the image of a brand, and such image processing has an important position in online consumption [22]. In addition, from the consumer's point of view, the choice of color is more than 62% involved, which not only distinguishes products from competitors, but also affects mood, feeling, positivity, and negativity, which plays a role in forming an attitude toward consumer products [23]. However, colorful images do not always bring good marketing results, and depending on the situation, they may give a more intense feeling when using achromatic colors [24].

As such, the atmosphere of the product can be mainly confirmed by the color provided through the image. However, even certain opinions, such as consumer reviews made up of text or explanations from providers, can contain colors that represent the mood of consumer goods [25-27]. Consumers usually use a lot of reviews left by actual users about consumer goods, and this analysis has been proven through many studies. In related analysis, it was analyzed using natural language processing of artificial intelligence, and as a result of research, it is said that the evaluation according to the actual use of the product and the evaluation desired by the company show accuracy of more than 90%. In this way, depending on the text and evaluation results of the review, it is involved in direct evaluation after the use of consumer goods, but it also affects the actual consumer goods themselves and the image of related companies [28].

2.5 Techniques Related To Sentiment

Recently, through technology, we are trying to understand the emotional context behind it by using various data expressed by humans. This method, referred to as emotion analysis, can use customer feedback to check the assessment of the company or to see how satisfied it is with the product or service. In order to acquire related data, social media such as SNS with the most various expressions are used a lot, or used directly by the social media company. As a representative example, researchers at Microsoft Research Lab analyzed Twitter posts to predict which women are at risk of postpartum depression [29]. This was inferred by performing natural language processing on text, and was performed based on what data the computer understood the emotional context behind a person's words and left behind before the symptoms occurred.

Sensibility Ergonomics is applied not only to general analysis through data but also to strategies for sales. Sensibility Ergonomics means realizing a person's desire or image that stimulates a person's emotions with a specific product design. Typically, it is widely applied to the product image used for selling goods, and the product image already reflects the sensibility ergonomics aspect of the seller or company selling the product, so analyzing it has a significant value.

3. Program design

The program is designed with the following contents. Functions related to existing data collection are performed through crawling or scraping. It is used only for analysis through algorithms in the program and is designed so that additional extraction is impossible. And it is limited to online shopping sites that people usually use. Websites with poor utilization are not reliable, and it is not easy to reflect them in crawling code because HTML structures are all different. Basically, crawling data is not directly stored, and is used for image transmission and collection. The use of images shows the top 10 detailed description images of consumer goods at once so that some users can use

them, and the color sense of the image is expressed through histogram analysis. If necessary, detailed analysis of the product is performed, the atmosphere is analyzed through natural language processing using data from some top reviews and consumer product descriptions to infer the nature and quality of the product. For a simple test, a program that performs some functions was created and tested, and the code was python and OpenCV, pytorch, and selenium were used as the library. The overall configuration of the test performance design is as follows.

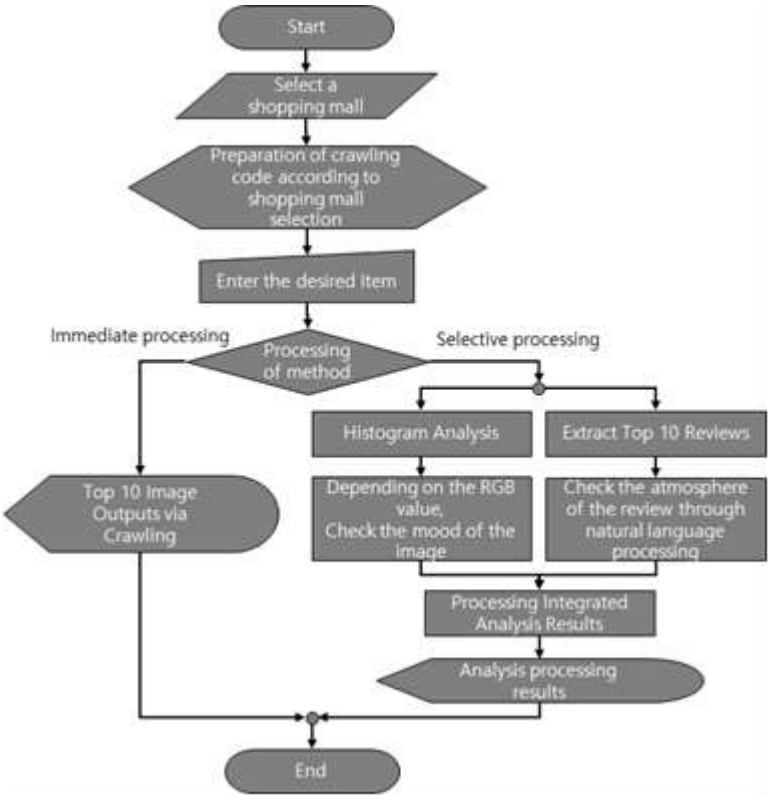

Fig. 3. Flowchart of designed program.

Figure 3 shows the flow chart of the designed program. The flowchart includes all the functions described above, and most of the general functions show the degree to which the person in charge of marketing can review them. Immediate processing by the program shows detailed images in the top 10 sales links of the product at once for comparison, and histogram analysis and review analysis, which lead to selective processing, are functions that can be confirmed through additional processing.

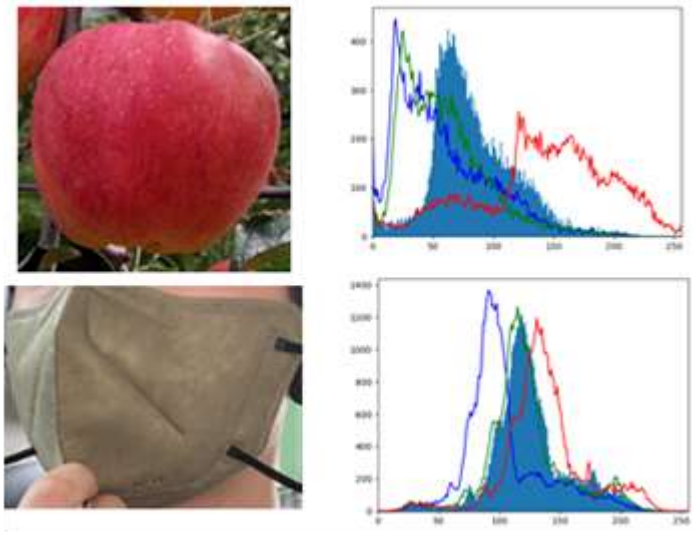

Fig. 4. Color analysis of goods image by histogram

Figure 4 is shown to explain the method of analyzing the image shown through the histogram. In general, when purchasing online, you purchase the product by looking at a picture of the product. The ripe apples are red and sell well. The image of the product also stands out as an apple with a lot of red colors. In addition, in the case of masks, it can be seen that masks with soft colors are released because they do not buy very strong ones except for achromatic colors.

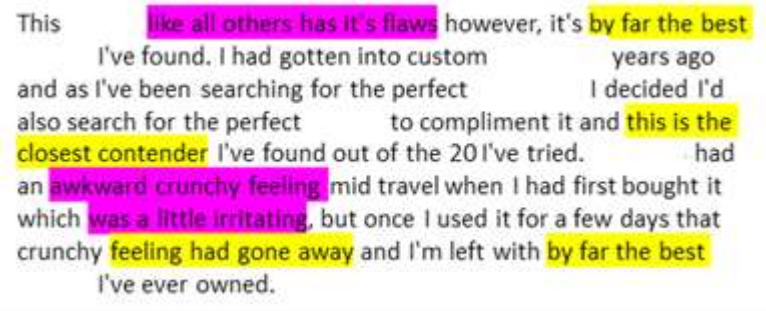

Fig. 5. Satisfaction analysis through product user review

Figure 5 shows a user review of a random object. The content of these reviews is analyzed through natural language processing, and the expression in yellow represents a positive expression and the expression in purple represents a negative expression. In this analysis, it is not judged by the number of expressions, but it is confirmed which is expressed more strongly. If you take this review as an example, it is a little inconvenient because 'by far the best' is mainly used, but it can be seen that the experience of using it was the best. With respect to Figures 4 and 5, each is valuable as individual data, but the data can be used to determine the integrated results. In general, customers want the product to be delivered in the same performance and form as described by the seller, so if the manufacturer meets the mood of the product, the purchasing layer, and the buyer's needs, customers who are willing to buy the product will pay without hesitation. Marketers can use the data generated by the corresponding process of the product to create the absolute value of the product they want to sell and the standard of the purchasing layer.

These functions simply seem to be performed. Although it is not technically difficult, there is no separate program to solve this problem in terms of marketing, and it takes a lot of time to solve only collection. The designed program may appear to simply output the top 10 data and compare and analyze it for marketing, but this data utilization is a very necessary part.

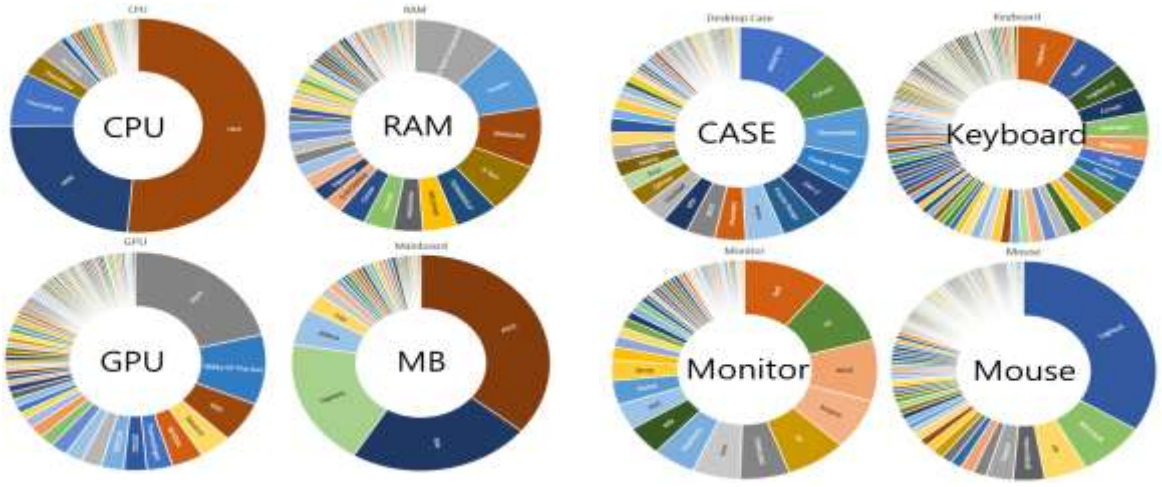

Fig. 6. Graph of 500 manufacturers of computer-related products by user evaluation at Amazon shopping mall on May 17, 2022.

Figure 6 is a graph of the top 500 data based on the number of reviews by computer parts on May 17, 2022 at Amazon, an online shopping mall. Although these are generally known facts, the following facts can be confirmed with the collected data.

- CPU is an oligopoly of Intel and AMD.
- The GPU market is sold by various companies, but ASUS is concentrated at the top.
- Regardless of the real share of the RAM market, products from various companies are being sold.
- Although keyboards are manufactured by various companies, there are certain companies that belong to the popular product line.
- Although mice are manufactured by various companies, Logitech's products are mainly used and evaluated well.
- The monitors are manufactured by various companies, and the top product groups are very evenly distributed except for sales.
- Mainboard is a market in which three to four companies form an oligopoly.
- Desktop cases are manufactured by various companies and sold in a wide variety of ways depending on the design that users need. There are quite a few expensive products in the top ranks, and the more expensive they are, the better the evaluation is.

In addition, the following details can be confirmed based on the above.

- Case where there are not many manufacturers regarding the product.
- A case where there are a lot of manufacturers in that product.
- A single product from a single manufacturer remains at the top.
- There are various manufacturers, but a single manufacturer in the top ranks has an oligopoly.

The above can be confirmed through simple crawling, and different opinions can be obtained in more detail depending on the person analyzing the data. Companies that want to start or approach a new business must perform this analysis on various products in order to succeed and go through the process of checking the so-called needs of users. Data with different meanings can be obtained if the product category is changed to be confirmed or crawled with different characteristics. It will take some time until a problem-free program is developed, but if used in the future, the time for marketers to search and grasp data will be drastically reduced, and financial losses from that time will also be reduced.

CONCLUSION

Previously, people searched and collected directly for marketing, and even if data is collected automatically, there is no separate program that provides analysis. In general, when collecting only some data that is needed, marketing personnel determine and use the accuracy of the data. When using a large amount of statistical data, in addition to direct collection, data may be purchased by paying an online sales company, or data surveyed by research institutes may be used. Despite various methods and efforts, collecting data may not be the desired data or may not be utilized directly. In fact, many companies are putting time and money into marketing, which is an investment for success in the company, but also a part that can reduce the various costs invested.

In the case of the program through the design proposed in this paper, it can be said to be a program tool that can be used for data collection and analysis easily, and it can replace some tasks that take time. Although it is not possible to collect all the details of the desired data, it is conveniently performed through the program regarding the collection method that must be performed uniformly by human hands. In addition, evaluations regarding the purchase desire to see and feel detailed description images that can vary from person to person are combined with textual data such as reviews by generating characteristics of image data, and then performing natural language processing to provide the inference value of the product.

When comparing the method with the passive data collection method, data from various companies could be confirmed compared to data from research institutions mainly represented by large companies. In the case of computer parts, manufacturers are often limited and large companies sell well, but it was also confirmed that products based on "sell well" are at the top of the list due to detailed images of people, preferences of real users, and necessary reasons. Although the program through this paper does not introduce difficult techniques or methods, it is expected that the existing method of marketing can be changed according to the present day by improving the data collection method and the evaluation method of that data.

Acknowledgments

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the Grand Information Technology Research Center support program(IITP-2023-2016-0-00318) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation). And this work was supported by Dong-eui University Grant(202301460001).

REFERENCES

[1] S. Ugalmugle and R. Swain, “Protective Face Masks Market Size By Product” (2021)

[2] Mordor Intelligence, “Disposable face mask market-Growth, Trends, COVID-19 Impect, and Forecast(2022-2027)” (2021)

[3] Grand View Research, “Disposable Face Mask Market Size, Share & Trends Analysis Report By Product (Protective, Dust, Non-woven), By Application (Industrial, Personal), By Distribution Channel, By Region, And Segment Forecasts, 2022 – 2030” (2021)

[4] Precedence Research, “Disposable Face Mask Market (By Product: Dust, Protective, and Non-woven; By Application: Personal and Industrial; By Distribution Channel: Offline and Online) - Global Market Size, Trends Analysis, Segment Forecasts, Regional Outlook 2020 – 2027” (2021)

[5] C. L. Petra, “Digital Marketning Benchmarks Leveraged By Marketing Analytics Tools”, CroDiM, vol. 3, no. 1, (2020), 82-97.

[6] Bigcommerce, “Ecommerce 101: The History and Future of Online Shopping” (2021)

[7] Y. Jessica, “Global online sales reach nearly $4.29 trillion in 2020” (2021)

[8] A. V. Karin, “Global Ecommerce Forecast 2021” (2021)

[9] Fortune Business Insights, “E-Commerce Packaging Market to Reach $140.86 Billion by 2021-2028” (2022)

[10] eMarketer, “Retail Ecommerce Sales Worldwide, 2019-2024” (2020)

[11] Jam, F. A., Sheikh, R. A., Iqbal, H., Zaidi, B. H., Anis, Y., & Muzaffar, M. (2011). Combined effects of perception of politics and political skill on employee job outcomes. African Journal of Business Management, 5(23), 9896-9904.

[12] KOTRA, “U.S. e-commerce market trends” (2022)

[13] Grand View Research , “E-commerce Market Size, Share & Trends Analysis Report By Model Type (B2B, B2C), By Region (North America, Europe, APAC, Latin America, Middle East & Africa), And Segment Forecasts, 2020 - 2027” (2019)

[14] V. W. Douglas and M. A. Neil, “Benchmarking Marketing Capabilities for Sustainable Competitive Advantage”, Journal of Marketing, vol. 69, no. 1, (2005), 80-94.

[15] J. Babović, V. Raičević and M. Carić, “BENCHMARKING AS A FUNCTION OF COMPETITIVENESS AND EFFICIENCY IN BUSINESS”, Economics of Argriculture, vol. 59, no. 1, (2012), 115-127.

[16] Symanto psychology ai, “Benchmarking Data: Collection, Analysis & Interpretation” (2021)

[17] S. H. Hong, “An Implementation of Smart Price Tracker System Using Web crawling”, Master`s thesis, Seoul National University of Science and Technology (2015)

[18] H. Kasereka, “Importance of web scraping in e-commerce and e-marketing.”, SSRN Electronic Journal, vol. 1, (2020), 1-10.

[19] C. Dilmegani, “Web Scraping in e-Commerce: Use Cases & Challenges in 2022” (2022)

[20] D. Wei, S. Neel, P. Robinson, B. Anurag, “Is a picture really worth a thousand words?: - on the role of images in e-commerce”, Proceedings of the 7th ACM international conference on Web search and data mining, (2014), 633-642.

[21] Gray, “Why Product Images Are So Important (And How to Take a Good One)” (2022)

[22] Sadiku, K. M. ., Mjaku, G. ., & Qarri , A. . (2023). Consumer Behavior in Decision Making - What Are the Factors?. *International Journal of Membrane Science and Technology*, *10*(3), 1-6. https://doi.org/10.15379/ijmst.vi.1257

[23] B. Paul, “The interactive effects of colors and products on perceptions of brand logo appropriateness”, Marketing Theory, vol. 6, no. 1, (2006), 63-83.

[24] S. Satyendra, “ Impact of color on marketing”, Management Decision, vol. 44, no. 6, (2006), 783-789.

[25] L. Xin, W. Mengyue, C. Yubo, “The Impact of Product Photo on Online Consumer Purchase Intention: an Image-Processing Enabled Empirical Study”, Proceedings of the 19th Pacific Asia Conference on Information Systems, 325 (2014)

[26] H. J. Lee, “The Effect of Black-and-White versus Color Imagery on Consumer Behavior: A Construal Level Theory Approach”, A doctoral thesis, The Ohio State University, 2016.

[27] F. Xing and Z. Justin, “Sentiment analysis using product review data”, Journal of Big Data, vol. 2, no. 5, (2015), 1-14.

[28] P. Ashwini Patil and G. Shiwani, “A Review on Sentiment Analysis Approaches” (2021)

[29] T.Wanliang, W. Xinyu and X. Xinyu, “Sentiment Analysis for Amazon Reviews” (2018)

[30] A. S. M. Arwa, “Product Sentiment Analysis for Amazon Reviews”, International Journal of Computer Science & Information Technology, vol. 13, no. 3, (2021)

[31] M. D. Choudhury, S. Counts and E. Horvitz, “Predicting postpartum changes in emotion and behavior via social media”, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, (2013), 3267.

DOI: https://doi.org/10.15379/ijmst.v10i4.1872