2024 International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS)

User Behavior Data Mining and Analysis of E-Commerce Platforms: a Study Based on Big Data

Haifeng Wang
School of Business
Chongqing Vocational College of Transportation
Chongqing, China
whf0410@163.com

2024 International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS) | 979-8-3503-6066-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/IACIS61494.2024.10721709

***Abstract*—The purpose of this study is to use big data technology to deeply dig and analyze the user behavior of e-commerce platforms, so as to solve the problems existing in traditional research, such as limited sample size, single data source, limited analysis methods, insufficient timeliness, and inability to deal with unstructured data. Based on the user behavior data of e-commerce platform, this study adopts K-Means++, K-means and hierarchical clustering algorithm to perform cluster analysis, focusing on recall rate as an evaluation index. Experimental results show that K-Means++ performs well in the recall rate, reaching 83%, which has a significant advantage over K-Means (63%) and hierarchical clustering (45%). The study highlights the excellent performance of K-Means++ in the analysis of e-commerce user data, providing a reliable way to improve personalized recommendations and user experience. The research results have guiding significance for e-commerce platform data mining and user behavior analysis, and provide strong support for practical application.**

Keywords—E-commerce platform, user behavior, data mining, big data, analysis and research

I. Introduction

In the current e-commerce environment, with the rapid development of Internet technology and the wide application of big data technology, user behavior data on the platform has become a valuable resource. By analyzing these data in depth, we can not only better understand consumer purchasing habits and preferences, but also provide more accurate personalized recommendations for e-commerce platforms, thereby improving user experience and platform efficiency. Traditional research methods often struggle to comprehensively capture and analyze complex user behavior data due to limited sample size, single data sources, and insufficient analysis methods. Therefore, this study aims to apply advanced big data technology to explore and solve these problems.

This article conducts in-depth clustering analysis on user behavior data of e-commerce platforms using K-Means++, K-Means, and hierarchical clustering algorithms, with a focus on examining the performance of these algorithms in practical applications. Research has found that the K-Means++algorithm performs well in terms of callback rate and has significant advantages compared to other algorithms. This finding has important guiding significance for optimizing data mining and user behavior analysis on e-commerce platforms.

The structure of the article is as follows: Firstly, the introduction section outlines the research background and significance; Next is a review of relevant work, commenting on historical research and pointing out its limitations; Next, provide a detailed introduction to the data processing and analysis methods used in this study; Next is the experimental results and discussion, showcasing the performance and comparison of different algorithms; Finally, summarize the entire article and propose prospects for future research directions. I hope that through this structural arrangement, readers can have a clear understanding of the overall picture and in-depth insights of this study.

II. Related Works

In past studies, scholars have focused on all aspects of e-commerce user behavior. However, dealing with large-scale, high-dimensional data, as well as unstructured data, remains a challenge. Although previous studies have proposed some methods, for example, Big data-based research Melnyk V explores the impact of social norms on consumer behavior through an interdisciplinary meta-analysis. The effect of social norms on approval behavior was found to be stable over time and across cultures, but the effect on disapproval behavior grew progressively, especially strong in survival and traditional cultures. Factors such as communication and cost moderated this effect, with communication to specific organizations or intimate groups promoting social norm compliance. These results provide important insights for management practices and future research [1]. Naim A explored the importance of consumer behavior in marketing, including the impact of consumer decision making, determinants of demand and purchase behavior. The focus was on different aspects and dimensions of consumer behavior and the most effective types of customer segmentation [2]. Busalim A systematically reviewed the research on sustainable fashion consumer behavior, collecting 167 journal papers and synthesizing and analyzing 88 of them. It was found that the number of papers has increased since 2009, but there are fewer qualitative, experimental, cross-cultural and longitudinal studies. Big data techniques are under-researched and many studies lack adequate theoretical foundations. The gap between consumer attitudes and behaviors needs to be further explored [3].

Pauluzzo R explored whether Generation Y consumers exhibit socially responsible consumption behaviors in fast fashion and found that it depends on the mix of consumer values. The results suggest that different combinations of consumer values influence their behavior. This provides guidance to stakeholders in the fashion industry and helps to develop more effective policies and strategies [4]. Mariam S explored the impact of debt level, market orientation and financial literacy on consumer behavior and financial performance of microenterprises. The results show that good debt management, market orientation and financial literacy are critical in maintaining consumer relationships and improving financial performance. This emphasizes the importance of awareness, education and wise management in both business and consumer ecosystems, providing insights into improving financial performance and encouraging consumers to make more informed financial decisions [5].

979-8-3503-6066-0/24/$31.00 ©2024 IEEE

Waliuddin A N aims to understand the significance of business strategy implementation and consumer purchasing behaviors in order to increase firm value. Through the case study, it was found that CV. Homie Indonesia uses multiple business strategies and is influenced by cultural, social, personal and psychological factors. The results of the study emphasized the important role of MSMEs in the national economy, as well as their competitive advantage and adaptability in different fields [6]. Kamkankaew P aimed to clarify the concept, definition and importance of marketing anthropology in marketing management. By analyzing academic sources, he reveals the key role of marketing anthropologists in understanding customer behavior and improving the success of marketing strategies. The application of anthropology strengthens the understanding of cultural differences, consumer preferences, and globalization, and promotes cross-cultural communication and global engagement [7]. Leonard N A C argues that Indonesia is facing waste disposal challenges, and the Melati Waste Bank has made efforts to promote waste reduction activities. However, the lack of consumer interest in waste recycling products has resulted in the failure to achieve the goal. It was found that consumer purchase of waste recycling products depends on environmental knowledge, recycling behavior and government regulations [8].

Dang H L explored the factors affecting customers' purchasing decisions in the online environment, proposing key factors such as trust, attitude, satisfaction, and website design quality, and analyzing the data through structural equation modeling. The results showed that customers' purchasing decisions were positively influenced by these factors, including trust, attitude and satisfaction [9]. Sudirjo F explored the impact of Generation Z consumer behavior on website quality, privacy and security, shopping service and shopping enjoyment, and the mediating role of shopping enjoyment on online impulse shopping. Through quantitative research and structural equation modeling analysis, it was found that website functionality directly influences online impulse shopping behavior, while online shopping services have a lesser impact on impulse purchases. Website quality, privacy and security, and shopping services all had an impact on the buying experience, and shopping pleasure increased online impulse buying behavior [10]. Sri Nurchaini D analyzed the decision making behavior of consumers in buying fresh vegetables in modern markets. He surveyed 35 people in Chambey town square and original shopping center. SEM-PLS ((Structural Equation Modeling - Partial Least Squares)) method was used to analyze the data and the results showed that internal factors (education level, occupation, income level) and external factors (culture, family, environment, social class) had a positive effect on consumer behavior. Fresh vegetables as a moderating variable also had a significant effect on consumer behavior [11].

Putri S A Z aimed to analyze the factors influencing shopping cart abandonment behavior while shopping in Shopee marketplace. He used multiple linear regression analysis through a survey of 100 school students in Surabaya City. The results showed that factors such as emotional ambivalence, indecision, willingness to pay, choice overload, website comparison, transaction inconvenience, cost, and risk have a significant effect on shopping cart abandonment behavior. Among them, selection overload was a significant factor influencing shopping cart abandonment behavior [12]. Geurin A N summarized the academic research on social media and consumer behavior in sports, including topics such as consumer use, engagement, user segmentation, and user-generated content. Future research priorities should include qualitative research, descriptive analysis, Generation Z consumer behavior, social media and purchase behavior linkages, and expansion to regions outside North America and underrepresented groups, as well as the adoption of new theoretical frameworks [13]. Kim N R investigated the impact of hair salon selection criteria on consumer behavior and customer satisfaction. The results showed that competence and relationship factors had a positive effect on consumer behavior and service and price satisfaction, while price factors had a negative effect on service satisfaction. Therefore, hair salons should focus on improving hairdressers' competence and relational skills rather than focusing too much on price to increase customer satisfaction and motivation for promotional activities [14]. Czeczotko M used PRISMA (Preferred Reporting Items for Systematic Reviews and Meta- Analyses) method to analyze the international literature on consumer behavior towards private label products. Forty-four eligible peer-reviewed studies were searched through Scopus and Web of Science databases. Most studies were found to focus on dairy products as the subject of analysis for private label products, with the main non-healthy choice factors being price and value for money. Only a few studies considered health factors, focusing on the shift of private label products from low-cost to sustainable brands with quality and health added value [15].

However, these methods can face problems with efficiency and accuracy when dealing with large amounts of data. This paper will learn from and extend these methods, combined with big data analysis tools, to address the shortcomings of existing research methods and more comprehensively understand e-commerce user behavior.

III. Methods

A. Data preprocessing and Cleaning

1) Data preprocessing : Good data is essential for the subsequent analysis to be carried out smoothly and to ensure that the results are more objective and accurate. Therefore, the processing of the collected data before the formal analysis is a very important part.

TABLE I. User behavior data set

User_ID	Age	Gender	Region	Product_Category	View_Count	Purchase_Count	Click_Count
1	42	1	4	4	41	5	15
2	54	0	1	2	36	5	22
…	…	…	…	…	…	…	…
100	58	1	4	4	42	0	20

The data set in this paper is a simulated user behavior data set of the e-commerce platform, which is used for the experiment of user behavior data mining and analysis . The data contains basic information about users, such as age, gender, and their browsing, purchasing, and clicking behaviors on the platform. The data set details are shown in Table 1.

Through the basic statistical analysis, it is found that there are only a very small number of null values and time errors. Considering the large amount of data, deleting this part of data has very little impact on the whole, so we regard this part of data as dirty data and delete it . Figure 1 shows the data preprocessing process.

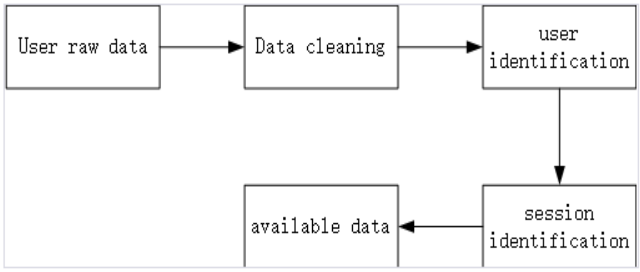

Fig. 1. Data preprocessing process

2) Data Cleaning: After obtaining the behavioral data of e-commerce users, we need to process the data, including removing redundant data or data with no research value. After deleting this data, we can perform operations such as user authentication, session identification, and historical order processing. This ensures that we get the right data. Data clearing is the first step of data preprocessing. By clearing data, it can filter out unnecessary data to further improve subsequent operations. The data to be deleted includes:

(1) Non-user login data, because our research object is the user, so non-user data is meaningless.

(2) Data that cannot open web pages or orders are closed due to network or server problems, because these data will increase the difficulty of research.

(3) Customer service data in e-commerce platforms, including communication records between robots and consumers, are ignored in order to obtain accurate data. These chat records are not within the scope of our research.

B. Feature Extraction and Analysis

Browsing goods on e-commerce platforms does not equate to users having purchased these goods, and such information is usually not directly available from the operator's data. However, studying and analyzing the behavior characteristics of users during browsing is a specific field of study. If we can predict the high consumption tendency of users, we can extract the characteristic attributes of their browsing behavior to provide personalized services. The purpose of this paper is to predict whether users belong to a group with high consumption tendency for a certain kind of goods by considering product type, user geographical location, visit frequency, waiting time and history.

1) The user accesses the commodity category attributes: Since users have different consumption tendencies for different product categories, it is necessary to classify users. When users browse or search for products, they extract the product name or search term from the web page through the URL, and then perform word segmentation to extract related nouns. Word segmentation tools are used to classify these nouns and compare them with predefined categories to determine the category of goods or search terms to which they belong and record them.

The products are divided into four categories, namely, clothing, food, general merchandise, and digital, to facilitate subsequent analysis. The similarity between key words and default classification is used to determine the product category.

2) Access frequency: The user is not limited to browsing the same category of goods on one e-commerce website, but takes into account all e-commerce website browsing history covering the product . This paper studies the high and low consumption tendency of users under different commodity categories, divided by commodity categories. Under this category, the user's visit frequency is defined as the sum of the number of visits to that category, divided by the sum of the number of visits of the user under all categories. The number of visits to these items can come from various e-commerce sites.

$$\mathrm{f}_t = \sum_i PV / \sum_u PV, i \in u \tag{1}$$

u represents all categories of goods, and i represents one category of goods. $\sum_i PV$ represents the total number of times the user accessed Category i goods, and $\sum_u PV$ is the total number of times the user accessed all categories.

3) Residence time: When users like a product, they stick around longer; On the contrary, for products they are not interested in, they will quickly leave. An increase in the amount of time a user spends in an interface can mean that they are more likely to buy something. This paper mainly studies the purchasing tendency of all users in each product category, so it is necessary to sum the page stay time of all users who visit the same product category and calculate its average . Since Internet users randomly browse products and are not sure which category they will view, the operator records the user's Internet surfing history in order of time stamps. Therefore, the residence time needs to be calculated using formula 2.

$$\mathrm{t}_i = \sum_i \Delta TS / \sum_i PV, i \in u \tag{2}$$

ΔTSrepresents the difference in time between the two most recent records in the processed data: $\Delta \mathrm{TS} > \varepsilon$, ε is a small positive number, and $\sum_\mathrm{i} \mathrm{PV}$ represents the total number of times the user accessed the Category i item.

4) Geographical location of the user: Due to unbalanced regional economic development, the average income of well-developed areas is basically higher than that of other cities, and their consumption power is much higher than that of poorly developed areas, so they are more willing to browse products and consume .

5) User history access records: By analyzing the Cookie information, you can obtain information about the products that the user has viewed in the past and compare it with the products that the user has visited recently. If it belongs to the same category, it means that users are more inclined to buy the product.

C. Big Data-Related Technologies

Spark is a memory-based distributed computing framework, and the project is mainly written in Scala language [16]. When manipulating elastic distributed data sets, the Spark program can extract data as easily as from a local database. Because Spark mainly performs data computation through memory, compared with Hadoop's MapReduce engine, it can retain all the intermediate results required by the computation process in memory, thus significantly improving the efficiency of iterative work. Compared with MapReduce, MapReduce needs to process intermediate results on disks, which increases disk read and write pressure. Spark introduces the idea of directed acyclic graph and optimizes the program running logic through shared memory region, which improves the computational efficiency by ten to hundreds of times compared with MapReduce [17].

Spark uses Spark Streaming to process real-time data, while Spark Core replaces Hadoop MapReduce as a core component. Spark Streaming has been widely used in real-time data processing and other fields. By combining the benefits of the Spark engine, it provides a powerful and flexible solution for real-time data processing. Although Spark replaces the compute engine, it is still closely related to Hadoop.

D. Model Construction

In the standard K-Means algorithm, the initial center point is randomly selected, and there may be errors when the amount of data is large. Therefore, the improved K-Means++ algorithm is chosen in this project to optimize the iteration [18]. The main idea of K-Means++ algorithm is different from K-Means in the choice of initial cluster center. It selects the initial cluster center through the following steps:

1) Selecting the first clustering center: Randomly selecting a sample point as the first clustering center.

2) Selecting the remaining cluster center: For each sample point, calculating its shortest distance from the selected cluster center (i.e. the distance to the nearest cluster center). Then, the next cluster center is chosen in a probabilistic manner, with the probability proportional to the square of the distance.

Repeating Step 2: Repeating the process of selecting cluster centers until K cluster centers have been selected.

The improved K-Means++ algorithm helps to improve the stability and convergence speed of the algorithm by selecting more representative initial clustering center. The probability of calculating the cluster center is shown in formula 3.

$$\partial = \frac{\mathrm{D}(x)^2}{\sum_{x \in X} D(x)^2} \tag{3}$$

D(x) is the distance of each element from its nearest center. Through such initialization process, K-Means++ is more likely to select widely distributed points in the data set as the initial clustering center, thus improving the effectiveness of the algorithm. After the improved initialization phase, the next K-Means iterative steps usually converge to the global optimal solution more quickly.

IV. Results and Discussion

The mathematical model used in this paper is as follows: Let D = {D_1, D_2, …, D_N} be the dataset containing user behavior data on the e-commerce platform, where each D_i represents a user interaction event such as browsing, clicking, adding to cart, or purchasing. Each interaction event is represented as a tuple (u, p, t, b), where u is the user ID, p is the product ID, t is the timestamp, and b is the behavior type. We define a user behavior matrix B ∈ {0, 1}^{M x N}, where M is the number of unique users and N is the number of unique products on the platform. Each entry B_{ij} in the matrix represents the user behavior towards product j by user i, where B_{ij} = 1 if user i has interacted with product j, and 0 otherwise. The evaluation of recommendation algorithm is closely related to product recommendation, which is usually carried out from two dimensions of user and item. The evaluation aims to measure the actual effect of the recommendation algorithm from multiple angles and find out the improvement points so as to optimize the recommendation service of the system in the future. When solving the Top-N recommendation problem, the commonly used evaluation indicators include Precision, F1 Score, and Recall.

A. Accuracy Evaluation

This experiment aims to compare and evaluate the performance and accuracy of three different Clustering algorithms, K-Means++, K-Means and Agglomerative Clustering, on simulated e-commerce user behavior data sets. Three clustering algorithms were used to cluster user behavior data, and the accuracy of each algorithm in the clustering process was calculated and compared with real clicks. The comparison results are shown in Figure 2.

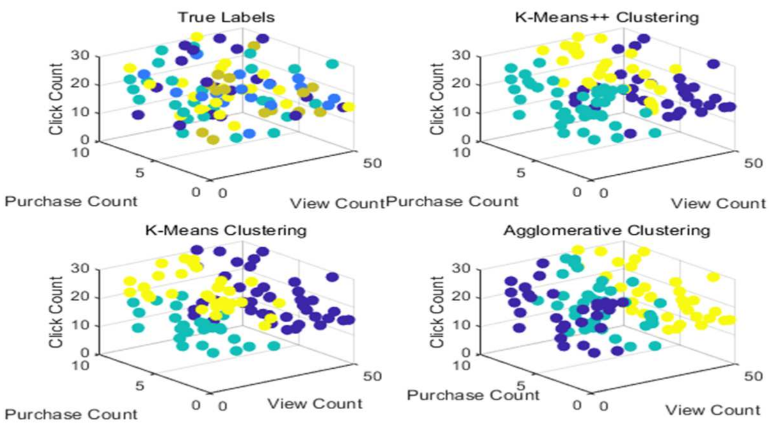

Fig. 2. Comparison of accuracy

Three different Clustering algorithms, K-Means++, K-means and Agglomerative Clustering, are applied to the e-commerce user behavior data set, and their accuracy is calculated. It can be seen from Figure 2 that the accuracy rate of K-Means++ reaches 80%, indicating that the algorithm completes the clustering task relatively well on the data set, helps to avoid falling into the local optimal solution, and improves the accuracy of clustering. The same clustering algorithm shows different performance characteristics on e-commerce user behavior data. It is important to select the clustering algorithm suitable for the actual scene and data distribution, and the evaluation of accuracy is helpful to intuitively understand the clustering effect of the model. Further experiments and parameter tuning can help optimize the performance of the algorithm.

B. F1 Value Evaluation

This experiment aims to evaluate and compare the performance of three different clustering algorithms on simulated e-commerce user behavior data sets, using F1 value as a performance evaluation indicator and comparing with real clicks. The result is shown in Figure 3:

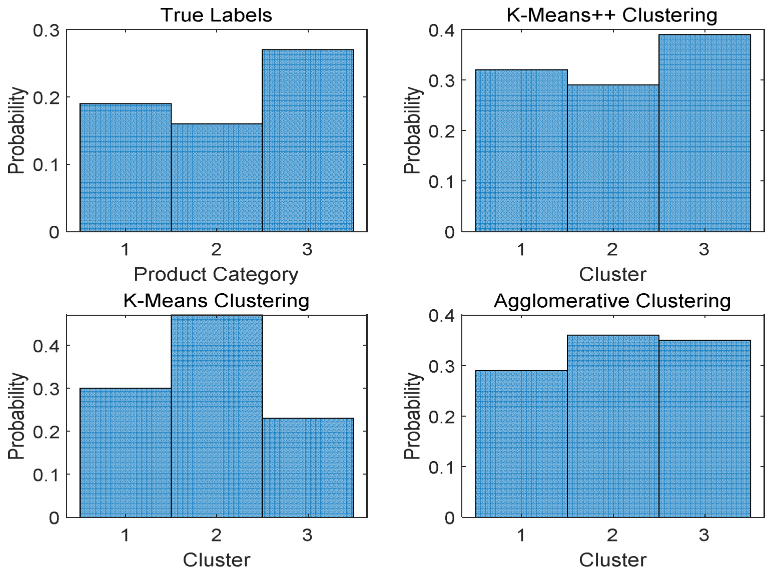

Fig. 3. Comparison of F1 values

As can be seen from Figure 3, F1 values of K-Means++ algorithm are higher than those of the other two algorithms. On the e-commerce user behavior data set, K-means ++ algorithm performs better on F1 value, F1 value is 0.31, and F1 value of K-Means algorithm is 0.24. The results show that the K-Means++ algorithm is more accurate and robust in processing such data, and has better performance and applicability. Choosing an appropriate clustering algorithm is crucial for accurate analysis of e-commerce user behavior, and K-Means++ performs well in this respect and can provide effective data mining and analysis means for related fields.

C. Recall Rate Evaluation

Through the simulated e-commerce user behavior data set, the performance of the three algorithms on recall rate is compared and evaluated, which provides a reference for selecting the suitable clustering algorithm. Three clustering algorithms are used to cluster user behavior data, and the recall rate of each algorithm is calculated and compared with the real one. The comparison results are shown in Figure 4:

As can be seen from Figure 4, the recall rate of K-Means++ algorithm is higher than the other two algorithms. On the e-commerce user behavior data set, the recall rate of K-means ++ algorithm is better, the recall rate is 83%, and the recall rate of K-Means algorithm is 63%. This shows that K-Means++ can more effectively capture the core characteristics of e-commerce platform user behavior, providing a strong support for accurate user analysis. Compared with other algorithms, the outstanding performance of K-Means-+ + in the key indicator of recall rate highlights its superiority in processing large-scale e-commerce user data, and provides a beneficial empirical basis for improving personalized recommendation and user experience on e-commerce platforms.

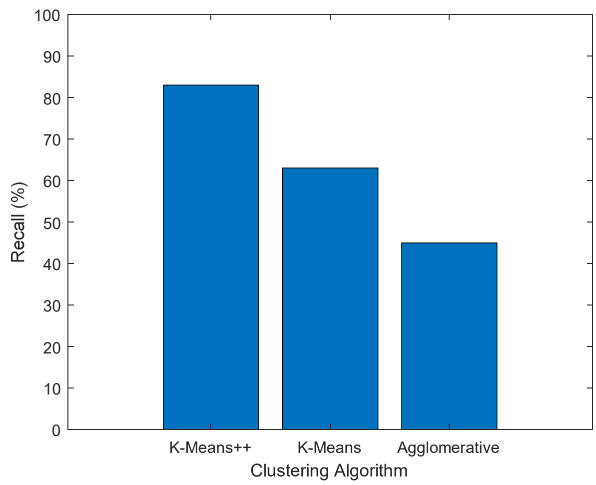

Fig. 4. Comparison of recall rates

V. Discussion

This research focuses on the data mining and analyzing of customer behaviour in a large data-driven e-commerce platform. The major results of this study are the identification of customer behaviour models, the creation of customized product recommendations, and the prediction of the sales trend. This knowledge will enable the EC to adapt its marketing tactics, improve the customer's experience, and improve the stock and pricing policy. Although there are some limitations to the use of historic data and data quality problems, we suggest that we invest more in the Data Analysis Facility to make effective use of user behaviour data. Unlike the conventional approach, Data Mining provides a more effective and cost effective approach to understanding the behaviour of users. All in all, our results highlight how important it is for Data Mining to grow and become more competitive in the digital market.

VI. CONCLUSION

Through in-depth mining and analysis of user behavior data on e-commerce platforms, this study successfully applied advanced clustering algorithms such as K-Means++ on the basis of big data technology. This paper highlights the importance of user behavior data mining in the field of e-commerce, especially emphasizes the application prospect of big data technology in this background. By comparing the experimental results, the K-Means++ algorithm performs well in the recall rate and provides a more accurate user classification and analysis tool for the e-commerce platform. In terms of experimental design, this paper fully considers the rationality of each step to ensure the reliability of the study. Future research directions include further refining the method details, expanding the experimental scale to increase credibility, and more clearly stating the substantial contribution of this study to the development of e-commerce platforms in the conclusion. This study provides useful experience and enlightenment for deepening the application of big data in the field of e-commerce.

REFERENCES

[1] Melnyk V, Carrillat F A, Melnyk V. The influence of social norms on consumer behavior: A meta-analysis[J]. Journal of Marketing, 2022, 86(3): 98-120.

[2] Naim A. Consumer behavior in marketing patterns, types, segmentation[J]. European Journal of Economics, Finance and Business Development, 2023, 1(1): 1-18.

[3] Busalim A, Fox G, Lynn T. Consumer behavior in sustainable fashion: A systematic literature review and future research agenda[J]. International Journal of Consumer Studies, 2022, 46(5): 1804-1828.

[4] Pauluzzo R, Mason M C. A multi-dimensional view of consumer value to explain socially-responsible consumer behavior: A fuzzy-set analysis of Generation Y's fast-fashion consumers[J]. Journal of Marketing Theory and Practice, 2022, 30(2): 191-212.

[5] Mariam S, Putra A H P K, Ramli A H, et al. Analysis of the Effect of Debt Level, Market Orientation, and Financial Literacy on Microenterprise Financial Performance: The Mediating Role of Consumer Behavior[J]. Atestasi: Jurnal Ilmiah Akuntansi, 2023, 6(2): 469-494.

[6] Waliuddin A N. Analysis of Business Strategy Implementation and Consumer Behavior In Purchase and Its Relationship To Company Value At CV. Homie Indonesia[J]. Jurnal Indonesia Sosial Sains, 2023, 4(03): 261-275.

[7] Kamkankaew P, Meesubthong C, Sawang K. Decoding, Connecting and Converting Cultural Understanding and Consumer Behavior: The Imperative of Applying Anthropology in Marketing Management[J]. International Journal of Sociologies and Anthropologies Science Reviews, 2023, 3(6): 1-26.

[8] Leonard N A C, Meilina R. Analysis of Green Consumer Behavior on Purchasing Waste Recycling Products Reviewed from Environmental Knowledges, Recycle Behavior, and Government Regulations[J]. MSJ: Majority Science Journal, 2024, 2(1): 207-218.

[9] Dang H L, Bao N V, Cho Y C. Consumer Behavior towards E-Commerce in the Post-COVID-19 Pandemic: Implications for Relationship Marketing and Environment[J]. Asian Journal of Business Environment, 2023, 13(1): 9-19.

[10] Sudirjo F, Lotte L N A, Sutaguna I N T, et al. The Influence of Generation Z Consumer Behavior on Purchase Motivation in E-Commerce Shoppe[J]. Profit: Jurnal Manajemen, Bisnis dan Akuntansi, 2023, 2(2): 110-126.

[11] Sri Nurchaini D, Nainggolan S, Saputra A. Analysis of the influence of internal and external factors on consumer behavior of fresh vegetables in the modern market in Jambi city[J]. GSC Advanced Research and Reviews (GSCARR), 2023, 15(2): 124-132.

[12] Putri S A Z, Rusdianto Y. Analysis of Factors that Influence Consumer Behavior towards Shopping Cart Abandonment in Online Shopping at Shopee[J]. East Asian Journal of Multidisciplinary Research, 2024, 3(2): 629-644.

[13] Rajan, P.T. and Ramesh, G.P., 2020. Mitigation of Power Quality in Wind DFIG-Fed Grid System. In Intelligent Computing in Engineering: Select Proceedings of RICE 2019 (pp. 615-624). Springer Singapore.

[14] Kim N R, Jin Y M. Effect of Hair Salon Selection Criteria on Consumer Behavior and Satisfaction[J]. Journal of the Korean Society of Cosmetology, 2022, 28(1): 1-9.

[15] Czeczotko M, Górska-Warsewicz H, Zaremba R. Health and Non-Health Determinants of Consumer Behavior toward Private Label Products—A Systematic Literature Review[J]. International Journal of Environmental Research and Public Health, 2022, 19(3): 1768.

[16] Liu Y, Lu J, Mao F, et al. The product quality risk assessment of e-commerce by machine learning algorithm on spark in big data environment[J]. Journal of Intelligent & Fuzzy Systems, 2019, 37(4): 4705-4715.

[17] Verma N, Malhotra D, Singh J. Big data analytics for retail industry using MapReduce-Apriori framework[J]. Journal of Management Analytics, 2020, 7(3): 424-442.

[18] Cui H, Niu S, Li K, et al. A k-means++ based user classification method for social e-commerce[J]. Intell. Autom. Soft Comput, 2021, 28(1): 277-291.