MODELLING AND ANALYSIS OF USER BEHAVIOUR

Brynza N. O., Koliesnik D.

Simon Kuznets Kharkiv National University of Economics, Ukraine.

E-mail: Natalia.Brynza@hneu.net,

Daniil.Koliesnik@hneu.net

Abstract

This paper provides a thorough study of modeling and analyzing user behavior in online shopping. The research is based on simulated interactions involving 40 users with an e-commerce platform while selecting different gaming monitors. Key behavioral indicators such as interaction time, specification views, and filter usage were standardized to allow for accurate clustering. Using hierarchical cluster analysis with Ward's method and Euclidean distance, three distinct user groups were identified, each demonstrating unique behavioral patterns toward BenQ, Asus, and Xiaomi products. Discriminant analysis was utilized to evaluate the classification's effectiveness and to identify the most significant variables affecting group differentiation. The findings showed an accuracy, confirming the reliability of the clustering method. This research emphasizes the significance of analyzing behavioral data to tailor user experiences, improve marketing tactics, and boost digital product usability. The findings offer important insights into how data-driven decision-making can enhance adaptive e-commerce solutions.

Intelligent data analysis goes beyond merely generating reports; it serves as a mechanism for continuous learning and improvement [1]. This allows organizations to adapt more effectively to changes in user behavior and market conditions. Utilizing data mining fosters a culture of data-driven decision-making, which is a crucial element for success in today's competitive environment.

In a world where data is becoming the new currency, intelligent analysis is crucial for its effective utilization. Applying this analysis can transform your understanding of users, leading to improved outcomes in your business.

User behavior refers to the totality of actions, interactions, decisions, and activity patterns exhibited by individuals when using a digital product or service. It is not merely a collection of random clicks or views; rather, it is a deliberate process that reflects a person's needs, intentions, interests, habits, and responses to the digital environment.

User behavior is fundamentally shaped by the dynamic interaction between individuals and digital systems. Analyzing user behavior is essential for various fields, including human-computer interaction, usability engineering, digital marketing, web analytics, and intelligent data analysis.

The data for this analysis consists of a sample that simulates the interactions of 40 users with an online store website as they choose among three different gaming monitors. The proposed approach not only classifies current behaviors but also allows for future scalability by adding new product items, such as an MSI monitor. This enables the study of user reactions using the developed classification model.

The following variables that characterize user behavior are essential for our analysis: interaction time, depth of study, use of site tools, and the data set contains information about the outcome of the interaction.

After completing the standardization procedure, all variables were converted to a uniform scale. This prevents any individual indicator from dominating the calculations of Euclidean distances in subsequent analyses. This uniformity is crucial for the accurate functioning of clustering methods, which rely on the distances between objects in multidimensional space.

Clustering is one of the key stages of analysis, allowing users to be grouped according to similarities in their behavior when choosing a product [2, 3]. This approach makes it possible to identify typical scenarios of interaction with the website, which can then be used to improve marketing strategies or personalize content. To implement cluster analysis in this study, a hierarchical method using Euclidean metrics and Ward's method was applied.

Based on user behavioral variables, a dendrogram was constructed using hierarchical cluster analysis by Ward's method and Euclidean distance as a similarity metric. Analysis of the dendrogram structure allowed us to identify three distinct clusters of users who demonstrate similar patterns of interaction with the online store interface (fig, 1).

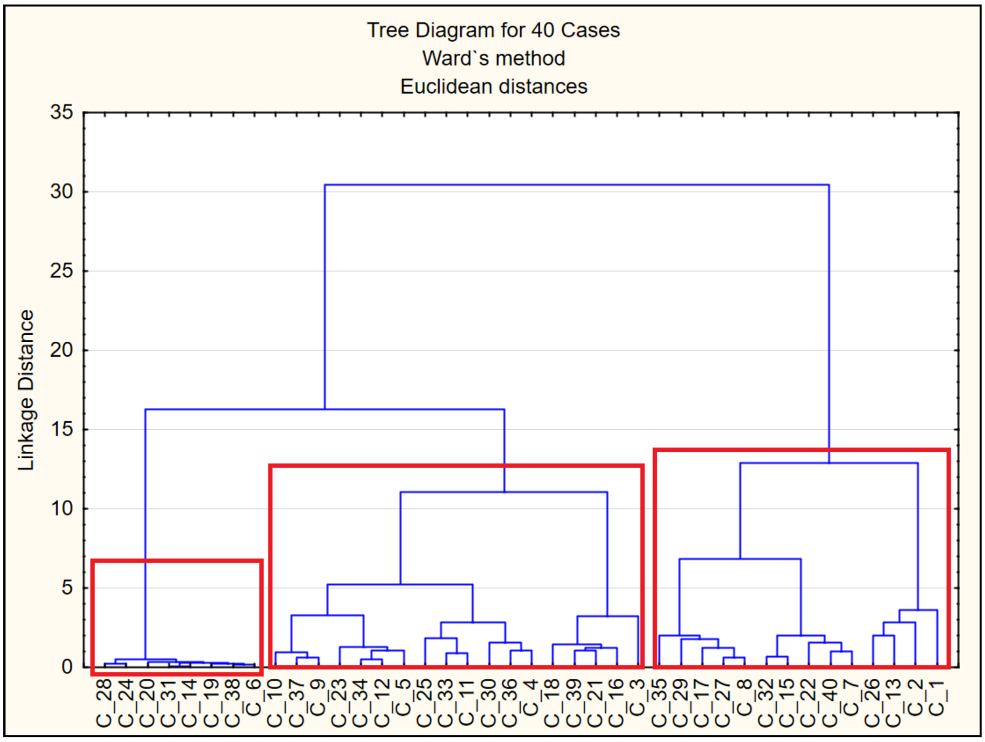

Fig. 1. Dendrogram of user clustering using the Ward method

Cluster 1 – users who interacted more frequently with the BenQ monitor, viewed its specifications, and used the price filter (fig. 2).

Cluster 2 – focused on Asus products and used the specifications filter.

Cluster 3 – focused on Xiaomi with high specification viewing and time on page values, indicating active interest.

Variable	Cluster No. 1	Cluster No. 2	Cluster No. 3
Time_BenQ	0,378195	-0,629294	-0,633160
Time_Asus	-0,715364	1,190531	1,197068
Time_Xiaomi	-0,426978	0,149039	2,258752
Specs_Views_BenQ	0,272457	-0,401255	-0,599406
Specs_Views_Asus	-0,574382	0,539571	2,106068
Specs_Views_Xiaomi	-0,400421	-0,028071	2,579830
Filter_Price_Used	0,480000	0,000000	0,000000
Filter_Spec_Used	0,000000	0,545455	1,000000

Fig. 2. Euclidean Distances between Clusters

The greater the distance, the more distinct the behavioral patterns of users in the respective groups are. According to the results obtained, users in clusters 1 and 2 are most similar in behavior, while cluster 3 is the most distant from the rest, indicating its unique behavior pattern associated with a high interest in Xiaomi products (fig. 3).

Cluster Number	Euclidean Distances between Clusters Distances below diagonal Squared distances above diagonal		
	No. 1	No. 2	**No. 3**
No. 1	0,000000	0,917576	3,743829
No. 2	0,957902	0,000000	1,743985
No. 3	1,934898	1,320600	0,000000

Fig. 3. Plot of Means for Each Cluster

The graph of average values for each of the three clusters (fig. 4) clearly shows differences in user behaviour. The most significant differences are observed for the variables Time_Asus, Time_Xiaomi, Specs_Views_Asus, and Specs_Views_Xiaomi. In the third cluster, the values of these indicators have the highest average values, which indicates the active interest of users in this segment in Asus and Xiaomi models. At the same time, the first cluster shows higher values for the Time_BenQ variable and more active use of the price filter.

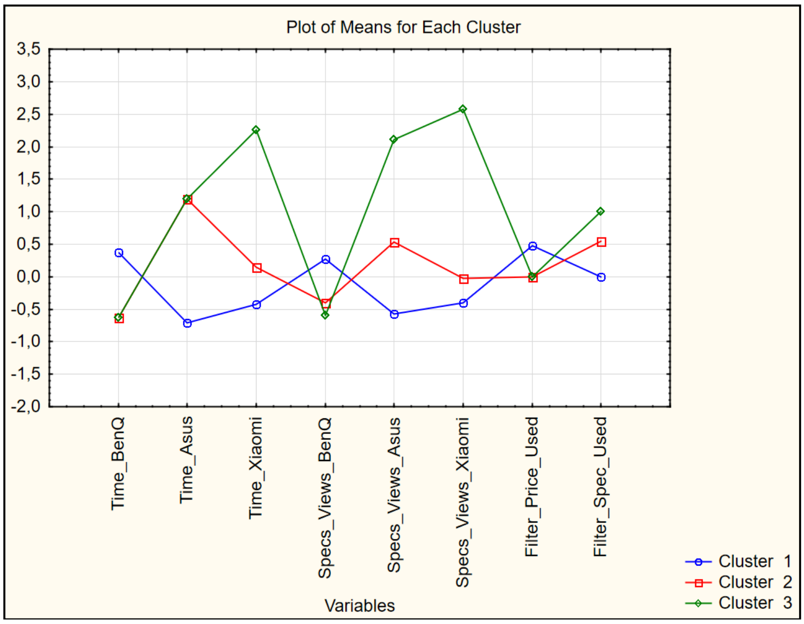

Fig. 4. Analysis of Variance

The purpose of using discriminant analysis is to evaluate how effectively user classes, defined through cluster analysis, can be distinguished based on certain statistical characteristics. These characteristics include the average viewing time for different monitor models, the number of views of technical specifications, and user activity in applying filters based on price and technical features [6].

Discriminant analysis will help identify the most significant characteristics for distinguishing between clusters 1, 2, and 3 [7]. It will also assess the accuracy of user classification based on these criteria. This analysis will enhance our understanding of which behavioral parameters can predict a user's class, which is crucial for personalizing marketing strategies and improving the overall user experience.

The implementation of discriminant analysis will help us determine if the current clusters are effective in distinguishing different types of users or if adjustments to the classification criteria are necessary. To achieve this, we will utilize data collected from our preliminary cluster analysis, which includes average viewing times for monitors of various brands (BenQ, Asus, Xiaomi), the number of views of technical specifications, and user activity related to filters based on price and specifications. This information will enable us to assess how well the user groups are differentiated and will support further analysis to identify the key features that influence product selection.

The results of the discriminant analysis are shown in fig. 5.

N=39	Wilks' Lambda	Partial Lambda	F-remove (2,29)	p-value	Toler.	1-Toler. (R-Sqr.)
Time_BenQ	0,012638	0,961329	0,58328	0,564478	0,220003	0,779998
Time_Asus	0,043265	0,280814	37,13568	0,000000	0,513220	0,486780
Time_Xiaomi	0,012964	0,937194	0,97172	0,390414	0,390023	0,609977
Specs_Views_BenQ	0,012397	0,980004	0,29585	0,746116	0,419574	0,580426
Specs_Views_Asus	0,023763	0,511284	13,85998	0,000060	0,578863	0,421137
Specs_Views_Xiaomi	0,021098	0,575869	10,67935	0,000335	0,546610	0,453391
Filter_Price_Used	0,012410	0,979025	0,31065	0,735381	0,437255	0,562745
Filter_Spec_Used	0,018984	0,639980	8,15697	0,001547	0,463125	0,536875

Fig. 5. Analysis of variables in the model

The factor loading matrix shows the correlation between each variable and the discriminant functions (fig. 6).

Variable	Root 1	Root 2
Time_BenQ	0,124365	-0,105383
Time_Asus	-0,543898	0,448283
Time_Xiaomi	-0,294773	-0,275796
Specs_Views_BenQ	0,093069	-0,057202
Specs_Views_Asus	-0,424353	-0,076299
Specs_Views_Xiaomi	-0,387444	-0,468484
Filter_Price_Used	0,131247	-0,113683
Filter_Spec_Used	-0,327894	-0,015802

Fig. 6. Factor loadings matrix

The significance of discriminant functions was assessed, variables that most influenced the distribution of observations between clusters were identified, and standard and factor coefficients were constructed for their interpretation. The resulting classification matrix showed 100% classification accuracy, confirming the high homogeneity of the formed clusters.

The most significant variables for discrimination: Time_Asus, Specs_Views_Asus, Specs_Views_Xiaomi, and Filter_Spec_Used. The results confirm the relevance of the behavioural characteristics used to divide users.

The Time_Asus variable has the most positive load on the second function, while Specs_Views_Xiaomi has a negative load on both, which allows us to interpret the direction of behavioural change between groups.

The results of the study confirm: the high quality of clustering and the validity of variable selection; the possibility of clear interpretation of online store user clusters; the identification of key behavioral indicators that can be used to adapt the interface, personalize recommendations, or target marketing.

References

1. Гороховатський В.О., Творошенко І.С. Методи інтелектуального аналізута оброблення даних: навч. посібник. – Харків: ХНУРЕ, 2021. – 92 с. **DOI:** https://doi.org/10.30837/978-966-659-298-2.

2. Brynza N. O. Development of a model for clustering countries of the world by the rate of demographic growth / N. O. Brynza, D. S. Lukianchikov // Інформаційні технології і автоматизація – 2024 : матеріали XVII міжнародної науково-практичної конференції, 31 жовтня - 1 листопада 2024 р. – Одеса: Вид. ОНТУ, 2024 р. – С. 30-32.

3. Tryon R.C. Cluster analysis. [Text] / R.C. Tryon L, D.E.Bailey. – McGraw-Hill Inc.,US – 1970.

4. Statistics and Analysis of Scientific Data / Massimiliano Bonamente. – New York : Springer Science+Business Media LLC, 2017. – 323 p.

5. Mining of Massive Datasets [Електронний ресурс]. – 2020. – Режим доступу до ресурсу: http://www.mmds.org/.

6. Evans M., Hastings N., Peacock B. Statistical Distribution. N.Y.: Wiley, 2000. Ch. 4. P. 31-33.

7. Бізнес-аналітика багатовимірних процесів : навчальний посібник [Електронний ресурс] /Т. С. Клебанова, Л. С. Гур'янова, Л. О. Чаговець та ін. – Харків : ХНЕУ ім. С. Кузнеця, 2024. Режим доступу до ресурсу: http://ebooks.git-elt.hneu.edu.ua/babap/index.html