Bulletin of the Transilvania University of Braşov
Series V: Economic Sciences • Vol. 17 (66) No. 2 - 2024
https://doi.org/10.31926/but.es.2024.17.66.2.3

EXPLORATION AND ANALYSIS OF AMAZON CUSTOMER BEHAVIOR - AN IT APPLICATION BASED ON CLUSTERING ALGORITHMS AND BUSINESS INTELLIGENCE SOLUTIONS

I. M. RUȚĂ[1] **D.A. LUPȘA-TĂTARU[2]**

Abstract: *The work focuses on exploring and analysing customer behaviour, identifying factors that have a significant impact on customers' level of satisfaction, and segmenting the customer sample into groups with similar characteristics using data mining, machine learning (ML), and business intelligence solutions. (BI). It presents the importance of approaching data analysis within the organization, as well as the usefulness of the results obtained for enhancing the company's performance.*

Key words: *behaviour analysis, satisfaction level, machine learning, business intelligence, Amazon*

1. Introduction

The accelerated evolution of e-commerce has led to the need for understanding customer behaviour for organizations whose goal is to increase their satisfaction and loyalty. E-commerce platforms generate large volumes of data, thus providing companies with the opportunity to implement methods for analysing customer behaviour. The data are the pulse of a company, and analysing them is fundamental in the process of making rational decisions based on real indicators.

Customer behaviour analysis is a strategic tool that allows companies to adapt their processes, enhance user experience, and optimize their marketing strategies. Nevertheless, businesses today are facing an onslaught of techniques and models dedicated to these analyses, yet few of them know how to leverage them. Companies need clear guidelines to implement effective data analysis solutions to efficiently identify customer habits, the factors that influence their satisfaction, and consequently, methods to utilize the results obtained. The objectives of the research work focus on exploring and analysing customer behaviour, identifying the factors that have a significant impact on their level of satisfaction, and dividing the customer sample into segments with the same

[1] Transilvania University of Braşov, iulia.ruta@unitbv.ro
[2] Transilvania University of Braşov, lupsad@unitbv.ro

or similar characteristics.

Analysing customer interactions and satisfaction levels through data mining and machine learning (ML) algorithms provides a data-driven approach to deciphering the factors that influence customer decisions. This granular understanding allows for more targeted strategies, personalized marketing, and improvements in service quality that directly address the needs and expectations of each customer segment. Therefore, the need for analysing customer behaviour is evident, and this research highlights ways to implement data analysis and interpret its results, as there has been a lack of documentation regarding data analysis solutions that serve both a technical department, which uses the data for further developments (such as the implementation of artificial intelligence (AI) algorithms), and a marketing or management department that develops the company's growth strategies.

2. Research in the field

In the specialized literature, there are numerous scientific articles that attest to the importance of using data analysis techniques for evaluating customer behaviour and their level of satisfaction. The need for the development of these analysis processes is based on the significant evolution of electronic commerce.

The article “An advanced intelligence system in customer online shopping behavior and satisfaction analysis,” authored by N. Moon, I. Talha, and I. Salehin in 2021, emphasises that, for the emerging growth of the online market, it is essential to understand online purchasing behaviour and customer satisfaction. Safety, trust, and product quality play an important role in customer satisfaction. The quality of the product, the price of the product compared to the local market, the return policy, and the timely delivery of the product are also essential elements of online shopping.

Customer retention has become problematic in the context of the evolution of e-commerce, as the existence of a developed market encourages customers to experiment as much as possible, to choose the lowest prices or the best deals, which leads to a constantly changing preference behaviour and induces a degree of business instability, while at the same time encouraging a decrease in product quality just to be more attractive. In the article "Customer retention through service quality and satisfaction: using hybrid SEM-neural network analysis approach" published in 2022, the authors frame the provision of a relevant experience for the consumer as a distinctive hallmark of marketing activities. Additionally, a positive and engaging customer experience delights consumers and leads to their loyalty or to the repurchase of the product/service by satisfied customers.

The evolution of e-commerce is concurrent with the development of data mining techniques, the concept of machine learning, business intelligence solutions, as well as algorithms based on artificial intelligence. The evolution of these information tools has consistently supported the improvement of business strategies for analysing the target market and segmenting it. Customer segmentation is a common strategy for grouping consumers with similar needs, characteristics, and behaviours. The article "A novel time series clustering method with fine-tuned support vector regression for customer

behaviour analysis," published in 2022 by authors H. Abbasimehr and F. Baghery, presents traditional segmentation approaches using certain descriptive variables, such as customer demographic attributes, as being insufficient. When demographic variables are not available or cannot be derived from the available data, the traditional segmentation approach cannot be used. The main challenge of the current literature regarding customer segmentation is that customers often adopt static segmentation approaches. The fundamental disadvantage of static segmentation models is that they treat customer segments, and their members as fixed entities, when in reality, customer behaviour is not consistent over time and it evolves. Furthermore, using a static customer segmentation technique is insufficient, as it leads to the loss of crucial trends and patterns in consumer behaviour. Thus, this work contributes to the literature dedicated to the use of customer segmentation strategies employing data mining and machine learning methods, based on variables that hold value over time, in a context where data analysis becomes a current necessity, almost reliant on real-time analyses to identify trends and varied changes in customer behaviour, with the aim of finding innovative solutions and benefiting from a high degree of accuracy in decision-making support.

3. Research methodology

The analysis conducted is a qualitative one, aimed at highlighting the utility of tools based on data mining and machine learning (ML) techniques, alongside business intelligence (BI) solutions that complete the digital framework in the field of data analysis. The data used in this paper are public data from the company Amazon, intended to serve as educational support for conducting data analysis. The dataset, Amazon Consumer Behaviour Dataset, published on the Kaggle platform, contains data representing consumer responses to a satisfaction survey aimed at analysing consumer behaviour in order to identify areas for improvement in the company's operations and products, but especially for the development of recommendation systems to personalize consumer experiences. The dataset contains twenty questions relevant for identifying customer habits and satisfaction levels. The number of customers who provided responses for this questionnaire is six hundred consumers.

4. Research results

4.1. Identifying the characteristics of the clients

The respondents' population comprises 54% (323 respondents) in the 18-30 years age range, followed by 24% (144 respondents) in the 31-40 age range. This distribution is expected, as these are the age ranges in which the population is much more open to online shopping, which is also reflected in the company's data.

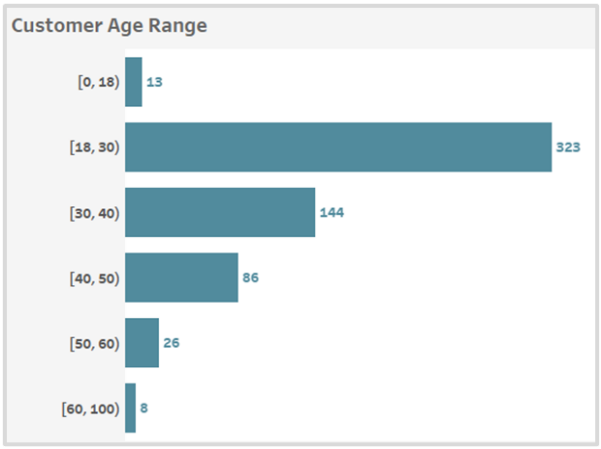

Fig. 1. *Age distribution of Amazon clients*

Regarding the respondents' gender, it can be observed that the distribution is not uniform for the analysed dataset. A 58% of the respondents (350) identify as female, while only about 24% (142 respondents) identify as male. This distribution confirms that female respondents exhibit more active shopping behaviour and are more open to providing feedback based on their purchases.

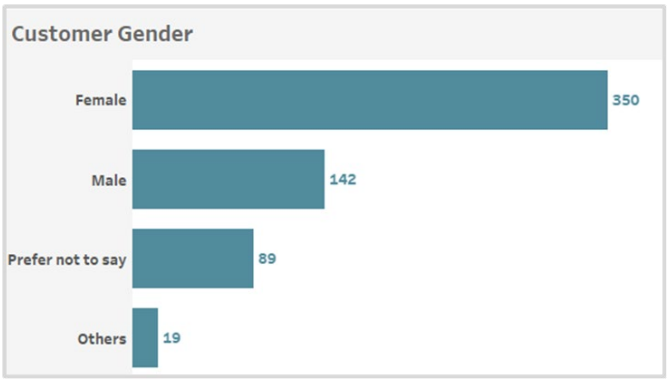

Fig. 2. *Age distribution of Amazon clients*

4.2. Analysis of variables with significant impact on the degree of satisfaction

The results of the analysis showed that the main factors influencing customer satisfaction according to the dataset are: the accuracy of the ratings, the importance given to reviews, the frequency of using personalized recommendations, the behaviour in using personalized recommendations, and the method of searching for products on the Amazon website and app.

To begin with, the variables that have a significant impact on the level of satisfaction were analysed, and relevant metrics were defined to measure customer satisfaction in

relation to the other variables. The data show that, following the analysis, the satisfaction level calculated based on the Net Promoter Score (NPS) key performance indicator is 29.50% (calculated as the difference between 48.67% promoters and 19.17% detractors).

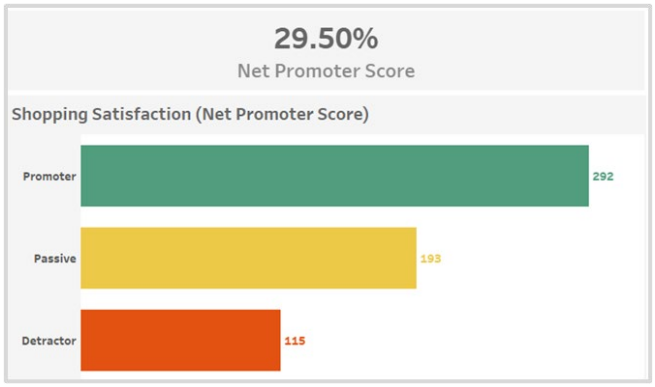

Fig. 3. *Net Promoter Score*

The media on the accuracy rating state that 48% of respondents rate the accuracy of the rating at 3, indicating that the scores given to Amazon products do not meet customer expectations and this area should see improvements. Furthermore, according to the distribution, between 19% and 36% of respondents do not assign a high level of importance to reviews. Thus, we can approach the hypothesis that the aforementioned accuracy of the rating is directly correlated with the importance given to reviews by respondents.

A percentage of 38% (228 respondents) make purchasing decisions for Amazon products based on personalized recommendations sometimes, while 41.67% (250 respondents) do not use this type of personalized recommendations at all.

A percentage of 45.33% of respondents state that they sometimes find personalized recommendations useful, while 28.50% of them claim that they do not. Thus, only 26.16% of respondents find personalized recommendations truly useful, a distribution that is also reflected in the frequency of purchasing the recommended products.

A percentage of 37.16% of respondents use categories for product search on the Amazon website or app, while 35.66% use keywords for searching, and only 21.16% of respondents use filters to find their preferred products.

4.3. Clustering analysis according to variables with significant impact on satisfaction level

The first perspective of clustering is based on the variable that measures the frequency with which respondents decide to purchase a product based on a personalized recommendation. Respondents who decided to purchase a product based on personalized recommendations have the lowest average satisfaction level, specifically 1.40 out of a maximum of 5. Thus, for these respondents, it can be stated that the recommendation system did not meet their expectations and needs improvements. One

of the hypotheses considered for this case could be the lack of relevant data for these respondents.

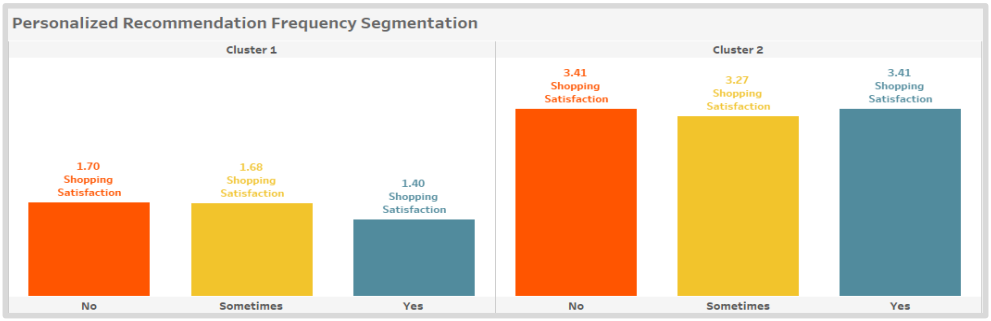

Fig. 4. *Clustering analysis for the variable measuring the frequency of use of personalized recommendations correlated with the respondents' level of satisfaction.*

The second perspective of clustering is represented by the utility that customers find in personalized recommendations. This variable determines that respondents who find these recommendations useful also have the highest average satisfaction level, specifically 3.50 out of a maximum of 5.

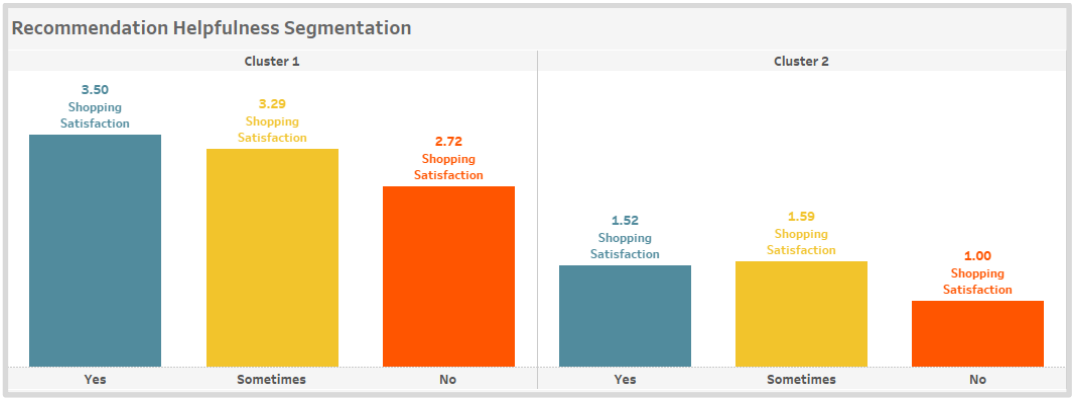

Fig. 5. *Clustering analysis for the variable measuring the utility of personalized recommendations correlated with the respondents' level of satisfaction.*

The third perspective of clustering presents the segmentation of respondents based on their product search behaviour on the Amazon website or app, viewed through the lens of their level of satisfaction. Although the majority of respondents stated that they most often use category selection to reach their desired products, the analysis from the clustering reveals that those respondents who utilize filters and keywords to search for products have significantly higher satisfaction. The average satisfaction level for respondents who use product categories to navigate to their desired products is 1.77 out of a maximum of 5.

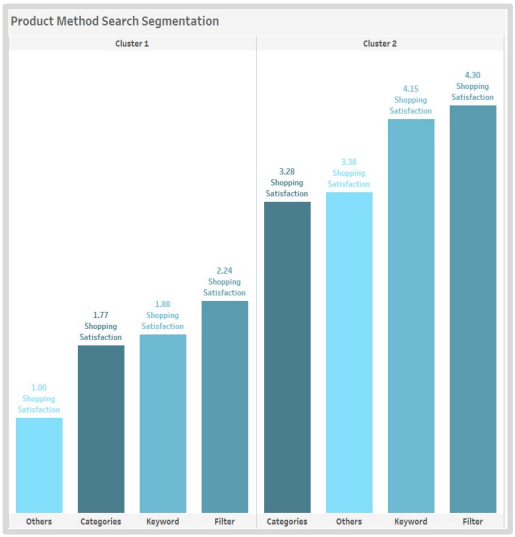

Fig 6. *Clustering analysis for the variable of product search method correlated with the respondents' level of satisfaction.*

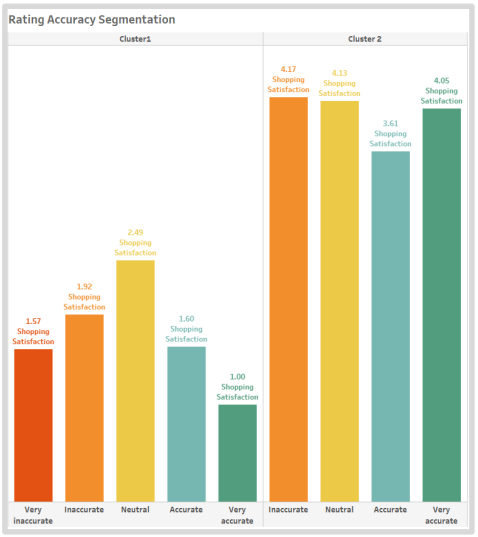

Fig 7. *Clustering analysis for the variable that measures the accuracy of the rating value correlated with the respondents' level of satisfaction*

Another perspective on clustering is dedicated to the accuracy of product rating values correlated with the respondents' level of satisfaction. Respondents who are completely dissatisfied with the accuracy of the rating have an average satisfaction level of 1.57 out of a maximum of 5 in the first segment.

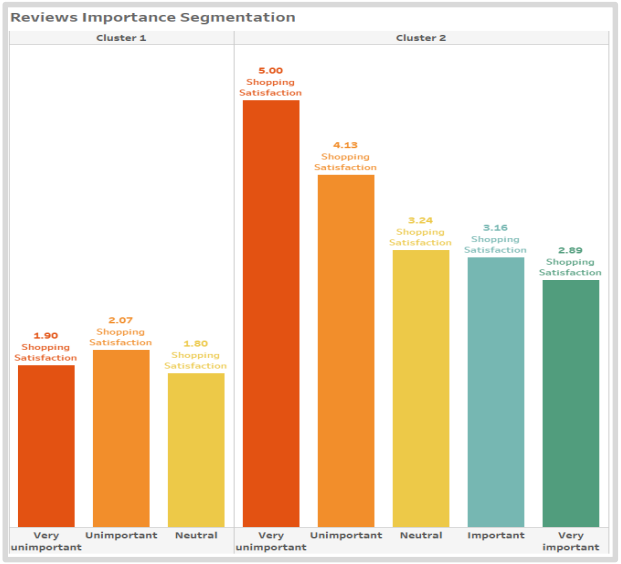

Fig 8. *Clustering analysis for the variable measuring the importance of reviews correlated with the respondents' satisfaction level.*

The last clustering perspective is dedicated to the importance of product reviews correlated with the respondents' level of satisfaction. Increased attention must be given to respondents for whom reviews hold a high degree of importance. Customers for whom reviews are important or even very important have an average satisfaction rating of 3.16 and 2.89 out of a maximum score of 5.

5. Discussions, limitations and conclusions

In a realistic framework, this approach can be used to analyse a company's data, being adapted according to the specifics of the data and the management requirements that will utilize the results of the analysis. In this paper, the practical applicability of advanced data analysis tools was simulated to create both general and detailed insights into customer behaviour and the factors influencing their level of satisfaction.

In this work, a small dataset was used, which optimized the performance of the digital tools employed to analyse the data. In a scenario where the volume of data is significantly larger, the performance of the data mining, machine learning (ML), and business intelligence (BI) tools used can be put to test. Additionally, in a context where the data used are characterized by numerous anomalies such as missing values, different data types, or the presence of outliers, the analysis conducted would require modifications to increase the level of data accuracy. In conclusion, the paper highlights the importance of adopting data analysis methods within a company that aims to achieve its long-term goals based on understanding customer behaviour, customers' characteristics, the factors that influence their habits, and their level of satisfaction. Thus, the implementation of such a solution can be the foundation for the company's strategic decisions aimed at enhancing its performance.

References

*A*bbasimehr, H., Baghery, F., 2022. A novel time series clustering method with fine-tuned support vector regression for customer behavior analysis. *Expert Systems with Applications,* Vol. 204. Elsevier.

Moon, N., Talha, I., Salehin, I., 2021. An advanced intelligence system in customer online shopping behavior and satisfaction analysis*. Current Research in Behavioral Sciences,* Vol. 2. Elsevier.

Otto, P.E., Davies, G.B., Chater, N., Stott, H., 2009. From spending to understanding: Analyzing customers by their spending behaviour. *Journal of Retailing and Consumer Services,* Vol. 16(1): 10-18. Elsevier.

Salamah, A., Hassan, S., Aljaafreh, A., Zabadi, W., AlQudah, M., Hayat, N., Mamun, A., Kanesan, T., 2022. Customer retention through service quality and satisfaction: using hybrid SEM-neural network analysis approach. *Heliyon,* 8(2): e10570.

https://www.kaggle.com/datasets/swathiunnikrishnan/amazon-consumer-behaviour-dataset