European Research Studies Journal
Volume XXVII, Special Issue 2, 2024

pp. 49-58

Analysis of Consumer Behavior Using an Intelligent Multi-Source System

Submitted 18/02/24, 1st revision 16/03/24, 2nd revision 20/04/24, accepted 16/05/24

Mariusz Mazurek[1], Magdalena Hałas[2], Jan Sikora[3],
Anna Wiśniewska - Vistula[4], Diana Wróblewska[5], Sebastian Zupok[6]

Abstract:

Purpose: *This work aims to develop an innovative system that analyzes multi-source data and human behavior, ultimately creating and sharing improved procedures and solutions. It focuses on building an IT system prototype for behavior analysis, optimizing the data mining process, and generating innovative business processes.*
Design/Methodology/Approach: *The application aims to optimize processes, analyze data, and reveal relationships between data and processes. Business models will be created using external data, data warehouses (such as ERP systems), and data from online resources (web mining). A process database will support computational intelligence algorithms, with an agent responsible for gathering online data. New data management methods were developed and implemented, while algorithms were designed for efficient web data searching. The system will leverage artificial neural networks, statistical and stochastic methods, fuzzy sets, genetic algorithms, and combinations to build an intelligent computing system.*
Findings: *The innovative system will contribute new data management methods and algorithms for web data searching and analysis. The algorithms will advance methods and concepts for capturing, transmitting, collecting, and extracting information while providing suitable data presentation formats.*
Practical Implications: *The insights from this system have the potential to revolutionize the way businesses identify and optimize new processes, generate innovative business models, and strengthen their decision-making. By comprehensively analyzing multi-source data, this system can inspire and motivate professionals in the field of data analysis and process optimization.*
Originality/Value: *This research is at the forefront of developing and implementing a system for analyzing multi-source data and human behavior. By using cutting-edge techniques such as artificial neural networks, statistical and stochastic methods, fuzzy sets, and genetic algorithms, this work provides an intelligent and robust framework for mining data and optimizing business processes, which is sure to intrigue and interest academic researchers, data analysts, and business professionals in the audience.*

[1]*Corresponding Author: Institute of Philosophy and Sociology of the Polish Academy of Sciences, e-mail: mariusz.mazurek@ifispan.edu.pl;*
[2]*WSEI University, Lublin, Poland, e-mail: Magdalena.Halas@wsei.lublin.pl;*
[3]*WSEI University, Lublin, Poland, e-mail: Jan.Sikora@wsei.lublin.pl;*
[4]*WSEI University, Lublin, Poland, e-mail: Anna.Wiśniewska-Vistula@wsei.lublin.pl;*
[5]*WSEI University, Lublin, Poland, e-mail: Diana.Wróblewska@wsei.lublin.pl;*
[6]*Wyższa Szkoła Biznesu - National Louis University, e-mail: sebastian.zupok_xl@wp.pl;*

Keywords: *Consumer behavior, Intelligent System, Multi-source Data, Agent-Based Modelling.*

JEL codes: *C45, C61, L11, D12, D83, L86.*

Paper type: *Research article.*

1. Introduction

In a market economy characterized by a high degree of competitiveness, the consumer is at the center of attention of companies operating on the market (Rymarczyk, Cieplak, Kłosowski, Rymarczyk, 2018; Polakowski, Filipowicz, Sikora, and Rymarczyk, 2008). Getting to know its needs, preferences, and factors influencing how it behaves is one of the essential tasks of every company.

Creating a "portrait" of a potential consumer allows the organization to adjust its products and services to consumer expectations and ensure success in a dynamically changing environment. This is not a new concept (Bonabeau, 2002). The classic definition of marketing already places the consumer in the center of attention, as he is the key to achieving economic benefits for the company.

The rediscovery of the concept of end-consumer-oriented enterprise operation in recent years is the result of an increase in the degree of competitiveness in many industries, an increase in environmental variability, and growing consumer awareness. The basis of marketing activities is knowledge about consumer behavior and the factors that determine it.

Its knowledge enables the selection of appropriate marketing tools that allow for more precise control of buyer behavior (Demirkan and Delen, 2013; Baker, Donthu, and Kumar, 2016; Anjam *et al.*, 2020).

Needs are met through various types of material goods and services. How a given need will be met causes the need to make choices and, consequently, decide how to finance the purchase of products to meet the needs (Norena-Chavez and Thalassinos, 2023; Nermend *et al.*, 2021).

To satisfy their needs, an individual will undertake various actions conditioned by external factors from the environment and internal factors resulting from the human condition. In this case, it can be said that the individual's reaction to emerging stimuli is mainly related to:

− specificity of needs determining the sequence of consumption goals and aspirations,

− the size and structure of current and future consumption,
− acquiring material resources for consumption (Borshchev and Filippov, 2004).

The diversity of consumption needs is revealed to a different extent in each consumer, and the accompanying conditions mean that consumer behavior is usually individual.

However, one individual's attitudes and reactions to specific stimuli will never be entirely identical to the attitudes and responses of another individual.

However, if specific behavioral criteria are adopted, relatively homogeneous groups of consumer behaviors can be identified. However, this does not detract from the fact that each individual, in a specific way, by the individual scale of preferences and the available purchasing power, decides to choose a product and service and make a purchase (Cearley, Burke, Searle, and Walker, 2018; Pan and Choi, 2014).

2. Description of the Data Set and Analyses Performed

The data set containing information about the customer includes such variables as customer ID, customer age, gender, income, education, size of the town he comes from, and profession.

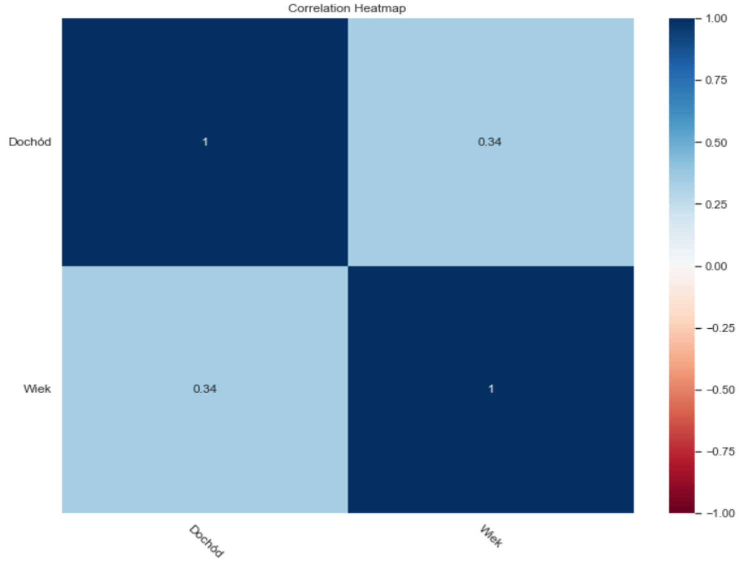
The profession variable is a categorical variable with the following levels: 0 - unemployed, unskilled, 1 - office worker, 2 - management, skilled worker, officer.

In the purchase data set, we have information about customer ID, day of purchase, purchase (0 - did not buy, 1 - bought), brand (0 - when the customer did not buy, 1 - first brand, 2 - second brand, 3 - third brand, 4 - fourth brand, 5 - fifth brand), number of pieces, last selected brand, number of purchased pieces for the previous brand chosen, price 1 (price for the product of the first brand), price 2 (price for the product of the second brand), price 3 (price for the product of the third brand), price 4 (price for the product of the fourth brand), price 5 (price for the product of the fifth brand), with active promotion for individual brands (Promotion 1, Promotion 2, Promotion 3, Promotion 4, Promotion 5), the client's gender, the client's income, the client's education, the size of the town he comes from and the client's profession.

Categorical variables were assigned states. You can check correlations between numeric variables. There is a weak correlation between age and income. Customers will be grouped using the k-means algorithm.

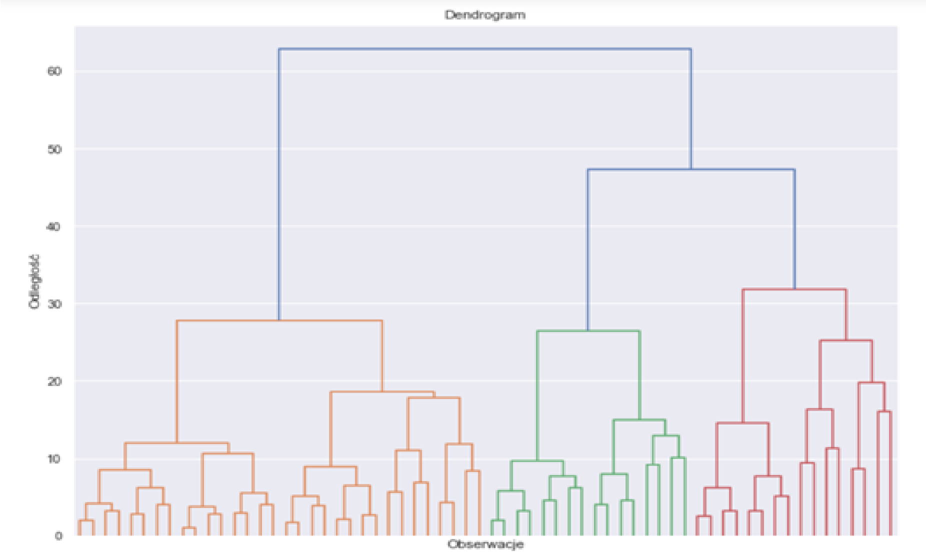
It is helpful to first build a hierarchical model to select the optimal number of clusters. Figure 2 shows a dendrogram representing hierarchical grouping using Ward's method. The optimal division is 2 or 4 clusters.

Figure 1. *Correlations between numerical variables*



Source: *Own creation.*

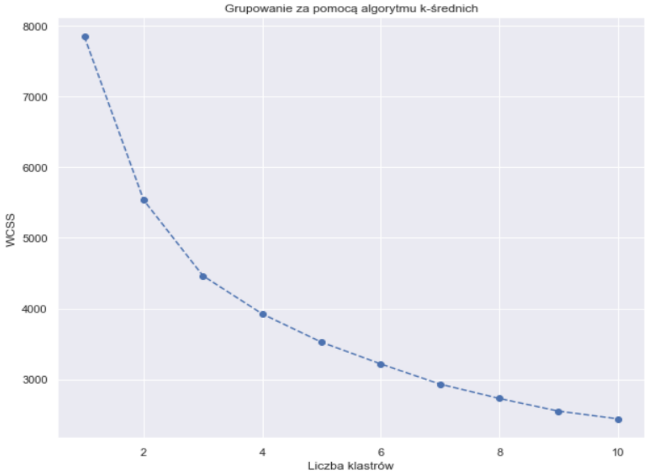
Figure 2*. Ward's hierarchical clustering model*



Source: *Own creation.*

Then, a k-means model was built with different numbers of k-means. From Figure 3, we can read that the optimal division is four groups.

Figure 3. *WCSS, depending on the number of clusters for the k-means algorithm*

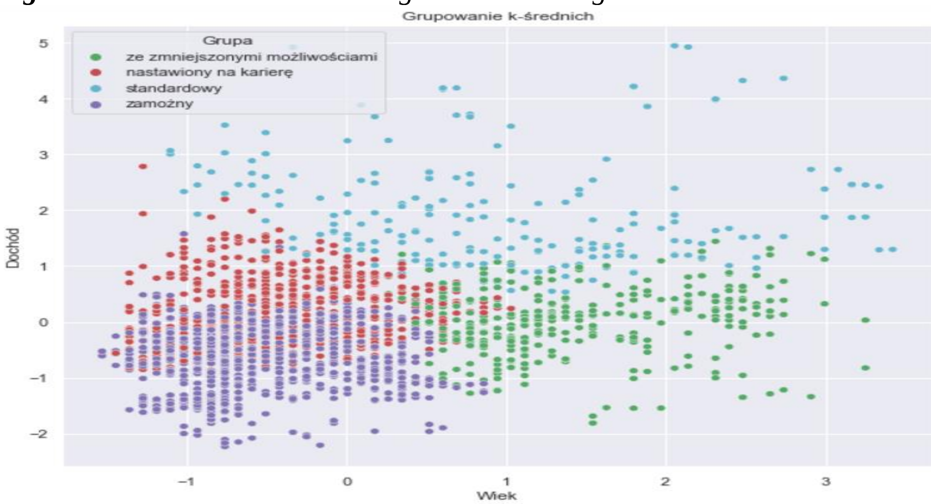


Source: *Own creation.*

Customers extracted using the k-means algorithm are shown in Figure 4. We can name the resulting groups based on the groups' characteristics. One of the groups of wealthy customers is older adults with higher education and high income. The second group is customers with reduced capabilities. These are mainly single, middle-aged, low-income people with low-level occupations, mostly in rural areas.

The group also includes a group of career-oriented clients. These clients are not single, are young, and have higher education. The remaining clients can be characterized as standard. These are customers with average income and secondary or higher education.

Figure 4. *Customers selected using the k-means algorithm*

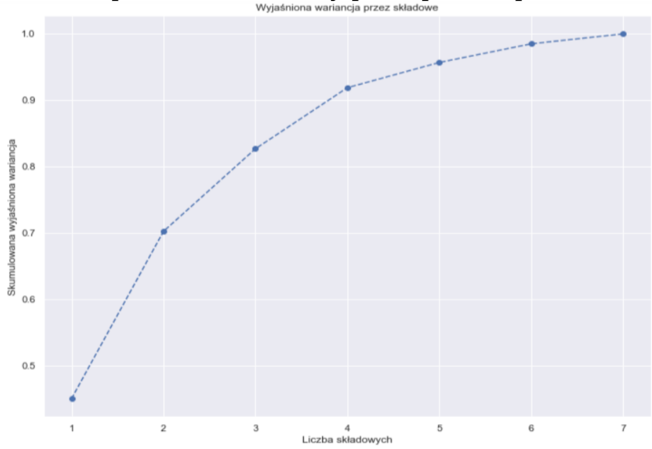


Source: *Own creation.*

The chart in Figure 4 shows the representation of customers. The group of career-oriented customers overlaps with the group of wealthy customers. We can only identify clusters of reduced and standard customers. To improve the clustering quality, we will perform clustering using PCA.

We plotted the cumulative variance explained by the appropriate number of components (Figure 5). Generally, we want to explain about 80% of the variance. Three main components achieve this level.

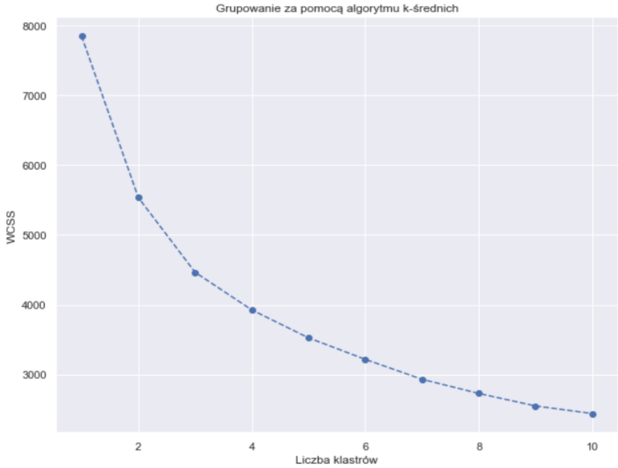
Figure 5. *Cumulative explained variance by principal components*



Source: *Own creation.*

After performing PCA, a k-means model with different numbers of k-means was built. From Figure 3, we can read that the optimal division is four groups.

Figure 6. *WCSS, depending on the number of clusters for the k-means algorithm*



Source: *Own creation.*

The graph in Figure 7 shows the representation of customers after grouping with the previous use of PCA. We see clearly marked customer groups (affluent customers, those with limited opportunities, career-oriented customers, and standard customers).

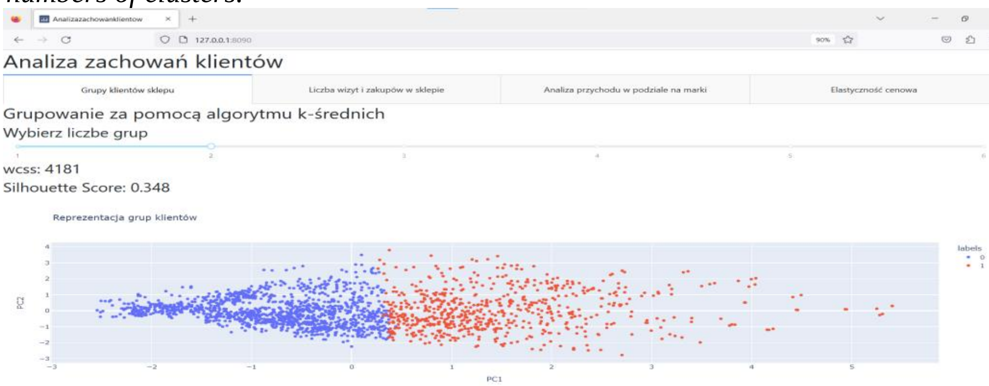
Figure 7. *Customers selected using the k-means algorithm with PCA*

Source: *Own creation.*

3. Building an Application Illustrating Customer Behavior

The presented analyses were presented as an application. The dashboard was created in Python using the Dash library. The user can choose one of four tabs: grouping of store customers, Number of visits and purchases in the store, revenue analysis by brand, and price flexibility, which are combined using DCC.Tabs components. After starting the application, the first tab, "Grouping store customers," is displayed in Figure 8.

Figure 8. *Application view after launch. K-means algorithm results for different numbers of clusters.*

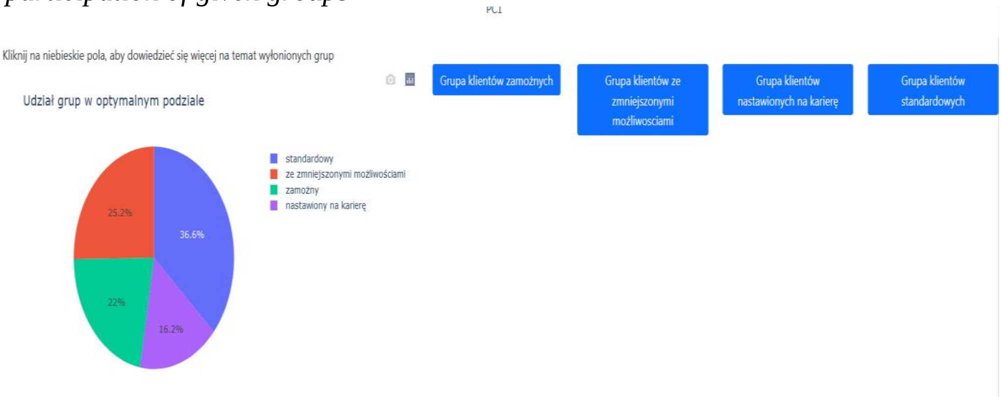


Source: *Own creation.*

The user sees WCSS and silhouette values for k-means clustering depending on the number of clusters selected. Using the slider, the user can choose the number of clusters from the following values, 1 - 6. Once the user selects the number of clusters, a representation of the customers will be displayed in the chart.

The k-means algorithm works on data previously subjected to principal component analysis with a selected number of principal components 3. Below the customer representation chart is a description of the optimal division (Figure 9). We see a pie chart showing the percentage of customers of a given group in the total number of store customers. The largest group are standard students (approximately 37%), and the smallest are affluent clients (approximately 22%).

Figure 9. *Description of optimal division. Pie chart of the percentage of participation of given groups*



Source: *Own creation.*

Figure 10. *Bar charts of average store visits and average purchases*

Source: *Own creation.*

The user receives information about obtaining more detailed information about the selected groups with the optimal number of groups. "Click the blue box to learn more about the selected groups." The second tab, "Number of visits and purchases in the store," displays a table presenting the average number of visits to the store, the average number of purchases in the store, and the ratio of the average number of visits to the store to the average number of purchases.

We also present bar charts showing the average number of visits to the store and the average number of purchases made, as well as box charts representing the spread of values for these features (Figure 10) divided into groups.

4. Conclusions

The diversity of consumer behavior results from the influence of external factors coming directly from the environment, i.e., the existing system of values, prevailing customs, traditional forms of behavior, and the mechanisms and institutions governing the modern economy. Each person's decisions are based on individual characteristics (attitudes, motives, personality). Therefore, it is challenging to capture and define all aspects and elements of consumer behavior on the market.

The methods of influencing the consumer, used by almost all sellers who want to actively participate in the e-commerce market and thus increase their profits, often take various forms, focusing their attention on multiple issues. Nevertheless, most techniques, such as price games, advertising messages, or free product delivery, attract customers and encourage them to purchase.

Future systems will be based on five critical technical areas: extensive data analysis, text analysis, web analysis, network analysis, and mobile analysis. User data is characterized as structured and user-generated online content rich in web information and unstructured and informal customer feedback. Identifying data sources opens the door to using various analytical techniques, including association rule mining, data segmentation, clustering, anomaly detection, graph mining, social network analysis, text and website analysis, and sentiment and emotion analysis.

Developing systems based on data from various sources and from many users enables the creation or re-innovation of social media monitoring systems, crowdsourcing data collection, social and virtual games, and more effective recommendation systems.

References:

Anjam, M., Khan, H., Ahmed, S., Thalassinos, E.I. 2020. The Antecedents of Consumer Eco-Friendly Vehicles Purchase Behavior in United Arab Emirates: The Roles of Perception, Personality Innovativeness and Sustainability. International Journal of Economics & Management, 14(3).

Baker, A.M., Donthu, N., Kumar, V. 2016. Investigating how word-of-mouth conversations about brands influence purchase and retransmission intentions. Journal of Marketing Research, 53(2), 225-239.

Bonabeau, E. 2002. Agent-based modeling: Methods and techniques for simulating human systems. Proceedings of the National Academy of Sciences, 99(3), 7280-7287.

Borshchev, A., Filippov, A. 2004. From system dynamics and discrete event to practical agent based modeling: reasons, techniques, tools. Proceedings of the 22nd international conference of the system dynamics society. Citeseer.

Cearley, D.W., Burke, B., Searle, S., Walker, M.J. 2018. Top 10 strategic technology trends for 2018. Retrieved from: https://www.gartner.com/doc/3811368/top–strategic-technology-trends.

Demirkan, H., Delen, D. 2013. Leveraging service-oriented decision support systems' capabilities: Putting analytics and big data in cloud. Decision Support Systems 55(1), 412-421.

Norena-Chavez, D., Thalassinos, E.I. 2023. Impact of big data analytics in project success: Mediating role of intellectual capital and knowledge sharing. Journal of Infrastructure, Policy and Development, 7(3).

Nermend, K., Łatuszyńska, M., Thalassinos, E.I. (Eds.). 2021. Decision-Making in Management: Methods and Behavioral Tools. Springer Nature.

Pan, A., Choi, T.M. 2014. An agent-based negotiation model on price and delivery date in a fashion supply chain. Production and Operations Management, 242(2), 529-557.

Polakowski, K., Filipowicz, S., Sikora, J., Rymarczyk, T. 2008. Tomography Technology Application for Workflows of Gases Monitoring in The Automotive Systems. Przegląd Elektrotechniczny, 84(12), 227-229.

Rymarczyk, T., Cieplak, T., Kłosowski, G., Rymarczyk, P. 2018. Design of Data Analysis Systems for Business Process Automation. IAPGOŚ 3, 43-46. DOI: 10.5604/01.3001.0012.5283.