ISSN 1507-3858
e-ISSN 2450-0003

Mateusz Bąska
Wroclaw University of Economics and Business
email: mateusz.baska@ue.wroc.pl
ORCID: 0000-0002-4465-7747

THE CLUSTERING AND SEGMENTATION OF CUSTOMERS AND PRODUCTS IN THE MULTI-CHANNEL SALES OF B2B E-COMMERCE TRADING COMPANIES

KLASTROWANIE I SEGMENTACJA KLIENTÓW I PRODUKTÓW W SPRZEDAŻY WIELOKANAŁOWEJ FIRM HANDLOWYCH DZIAŁAJĄCYCH W E-COMMERCE W SEGMENCIE B2B

DOI: 10.15611/ie.2021.2.02
JEL Classification: D80, C38, M30

© 2021 Mateusz Bąska
This work is licensed under the Creative Commons Attribution-ShareAlike 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by-sa/4.0/

Quote as: Bąska, M. (2021). The clustering and segmentation of customers and products in the multi-channel sales of B2B e-commerce trading companies. *Business Informatics*, (2).

Abstract: Given the progressive growth of e-commerce sales and the increasing interest in large-channel sales of business-to-business (B2B) trading companies among researchers and practitioners, the aim of this article was to identify the needs related to clustering and segmentation in B2B trading companies, as well as the techniques currently in use. Issues related to clustering and customer and product segmentation were explored and the most promising research areas for the nearest future identified. The article outlines the main techniques for clustering and segmenting customers and products, and identifies potential areas for further research. From a managerial perspective, the article is useful for companies entering the domain of multi-channel sales to guide their future efforts on methods to increase the value of customer purchases.

Keywords: clustering, segmentation, e-commerce, omnichannel, systematic review.

Streszczenie: Ze względu na postępujący wzrost sprzedaży e-commerce oraz rosnące zainteresowanie sprzedażą wielkokanałową firm handlowych w segmencie business-to-business (B2B) wśród naukowców i praktyków celem niniejszego opracowania jest przedstawienie aktualnego przeglądu literatury na temat możliwości zawansowanej analityki (Big data). W szczególności zbadane zostały zagadnienia związane z klastrowaniem i segmentacją

klientów i produktów, wskazano, jak temat ten był rozwijany w czasie, oraz zidentyfikowano najbardziej obiecujące obszary badawcze na najbliższą przyszłość. Artykuł oferuje wgląd w główne techniki klastrowania i segmentacji klientów i produktów, jak również wskazuje potencjalne obszary dalszych badań. Z perspektywy menedżerskiej artykuł jest przydatny dla firm wchodzących w sprzedaż wielkokanałową, aby ukierunkować ich przyszłe działania dotyczące metod zwiększania wartości zakupów klientów.

Słowa kluczowe: klastrowanie, segmentacja, e-commerce, omnichannel, przegląd literatury.

1. Introduction

The use of advanced analytics and machine learning by businesses is becoming increasingly common. The tools are becoming attractive to companies that operate in the business-to-business (B2B) segment and do not have as large a volume of customers and transactions as the business-to-customer (B2C) market. In this market segment, clustering and customer and product segmentation techniques are gaining popularity.

This article is a continuation of a study on the use of data analytics in trade enterprises dealing with online sales (e-commerce). The literature research shows that marketing uses data analysis in areas such as identification of customer requirements and market segmentation, solving many problems that would be very time-consuming or sometimes impossible to solve without using data analysis techniques and concepts (Bąska, Dudycz, & Pondel, 2019). To sum up, it can be concluded that companies in the retail area use data analytics to support decision-making processes.

Companies operating in the B2B segment have a different customer structure to those from the B2C segment. Customers in this segment are not anonymous, they are other companies, and therefore this gives one access to business profile information. As a result of the COVID-19 pandemic, many companies have decided to develop new sales channels by transitioning to a multi-channel or omnichannel model. One of the most important factors for these companies, after embracing new sales channels, is customer clustering and segmentation (Unity Group, 2021).

The main objective of the article was to identify the clustering and segmentation needs in trading companies that operate in the B2B sector and the relevant techniques currently in use. The result of the work is the identification of a preliminary research gap in the area of the clustering and segmentation of customers and products in the multi-channel sales of trading companies operating in the B2B e-commerce segment.

The study was divided into two stages. In the first stage, using the systematic literature review technique, an analysis of machine learning techniques and research directions related to clustering and segmentation was conducted. The aim of the literature analysis was to identify and organise the terminology related to clustering

and segmentation in decision-making processes in retail companies. In the second stage, the case study research technique was used to investigate the feasibility of using clustering and segmentation techniques in the B2B segment to support retail companies' decision-making processes.

Research questions were formulated for each stage of the study. In the first stage of the research, the following research questions were posed:

- (RQ 1.1) What machine-learning techniques are used for clustering and segmentation relating to e-commerce sales in the B2B segment?
- (RQ 1.2) What are the research directions related to the clustering and segmentation of customers and products in trade companies operating under the omnichannel model?

As far as the second stage of the study is concerned, answers were sought to the following research questions:

- (RQ 2.1) What are the main issues related to customer and product clustering and segmentation in trade companies operating under the omnichannel model?
- (RQ 2.2) What are the clustering and segmentation needs of companies regarding e-commerce sales in the B2B segment?

The article is divided into the following sections: the first part introduces the research context and the rationale for undertaking the study. The next section describes the first stage of the study – a systematic literature review together with the implementation of the research method on the basis of the case under analysis. The results of the research are presented. Section 3 presents stage two of the research and discusses the results of the applied case study method for two companies. A comprehensive analysis of the results obtained from stages one and two is presented in Section $, in which the conclusions from the study and further plans for work with respect to the subject of clustering and segmentation of customers and products in the multi-channel sales of trade companies operating in the B2B e-commerce segment are presented.

2. Literature review – stage 1 of the study

2.1. Research methodology

A systematic review is a defined and methodical way of identifying, evaluating, and analysing standardised primary research to investigate a specific research question. A systematic review can also reveal the structure and patterns of the existing research and therefore identify the gaps that can be filled by research (Kitchenham, 2004). Systematic reviews differ from ordinary literature studies in their formal planning and methodical execution. A good systematic review should be able to be replicated independently and will therefore have a different kind of scientific value than a simple literature study. When identifying, assessing, and summarising all available evidence relating to a particular research question, a systematic review can provide a higher level of accuracy in its findings.

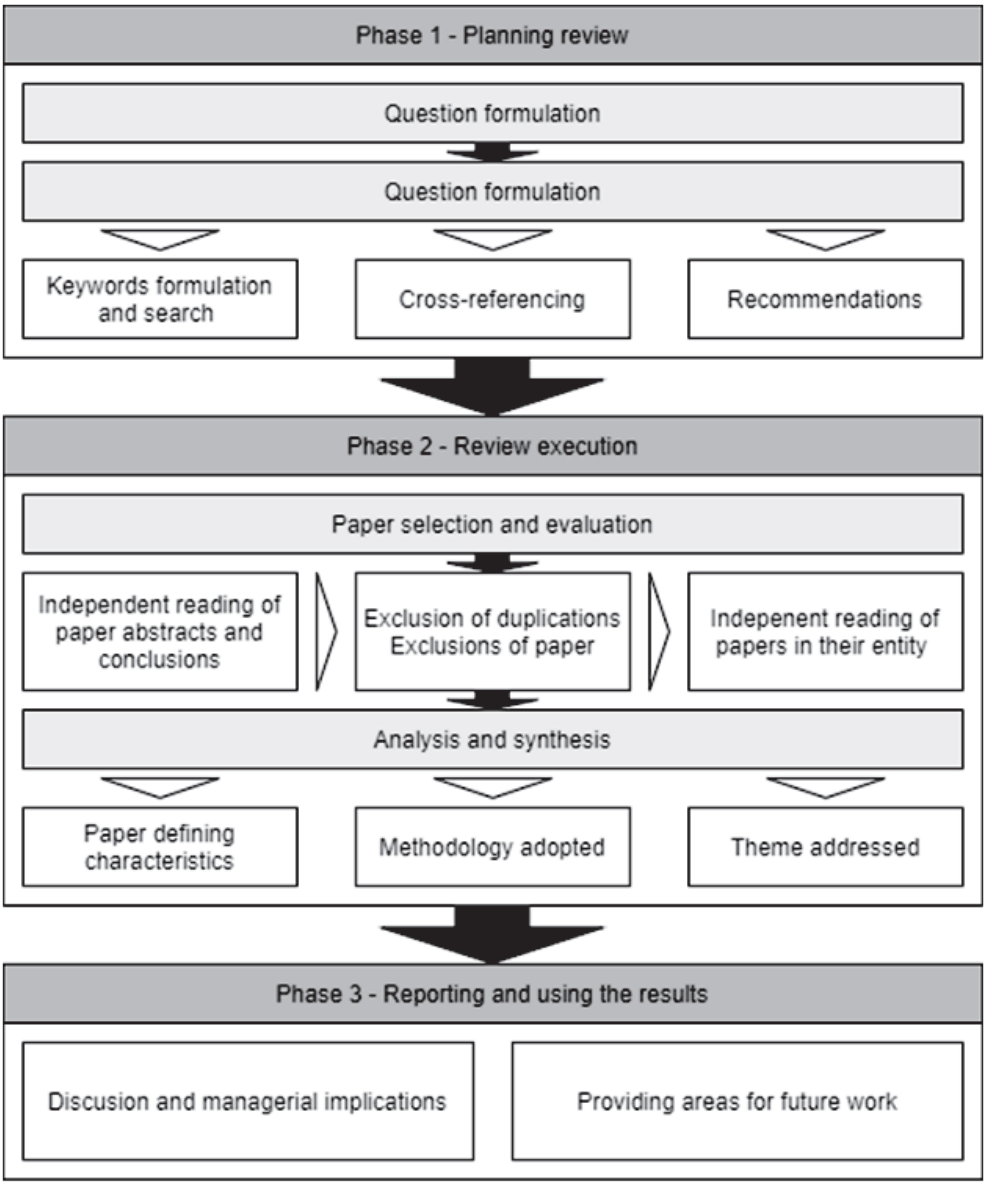

Fig. 1. Systematic literature review process

Source: (Denyer & Tranfield, 2009; Hofmann & Bosshard, 2017; Kitchenham, 2004).

Kitchenham (2004) identified the features that distinguish a systematic review from a traditional literature review, namely:

- defining and documenting an operational pattern of the research methodology for a systematic review prior to conducting the study to identify the research questions and activities to be applied in it;
- defining and documenting a search strategy within the research methodology to find as much relevant literature as possible with respect to the research area;
- description of the explicit inclusion and exclusion criteria within the methodology to be used to assess each result for further investigation;
- description of the quality assessment mechanisms within the methodology to evaluate each test result.

The study adopted the systematic literature review approach further described in (Bąska, Dudycz, & Pondel, 2019). This approach is an extension proposed by Kitchenham (2004) and further developed by Denyer and Tranfield (2009) as well as Hofmann and Bosshard (2017). The study identified three main phases of the systematic review process (Figure 1), i.e. review planning, review execution, and review reporting. The planning phase of the review should detail the stage of defining the research questions and the data sources to be searched for in the literature. The review phase should be divided into the stage of preliminary selection of obtained literature sources and their analysis, which should end with a synthesis (a list of the most important positions in in the literature in the area under study) of the obtained results in the context of the research questions posed.

All of the phases specified above have a defined sequence of steps, but the execution of the overall process involves the iteration, feedback, and refinement of the defined process. This section describes the three phases of the systematic review, and for their component phases, identifies certain important guidelines described by Kitchenham (2004).

A systematic review is characterised by clear procedures to find, evaluate, and synthesise the research findings related to the review question posed. The procedures were defined in advance to ensure transparency and repeatability of the systematic review process. This practice was also intended to minimise bias.

2.2. Execution of the study and results of stage one

Review planning

The first step in an in-depth literature review is to develop a clear research objective through clearly defined research questions, which need to be well-specified, informative, and clearly articulated to avoid ambiguity (Hohenstein, Feisel, Hartmann, & Giunipero, 2015). The following research questions were formulated:

- (RQ 1.1) What machine-learning techniques are used for clustering and segmentation relating to e-commerce sales in the B2B segment?

- (RQ 1.2) What are the research directions related to clustering and segmentation of customers and products in trade companies operating under the omnichannel model?

The first stage of the study defined the databases to be used for the literature analysis, i.e.:

- Scopus – http://scopus.com/.
- Web of Science – https://www.webofknowledge.com/.

A strategy was then developed to find scientific papers relevant to the research questions posed in stage one. The following strategy was used to build the search terms:

- Identification of keywords based on the research questions formulated.
- Verification of keywords in any publication relevant to the research topic.

The logical operators AND and OR were used in the search engine to build a search phrase, usable in databases that allow such a string with a logical operator.

Table 1. Search strings for database search (e.g. Scopus)

Scopus	(TITLE-ABS-KEY (clustering) OR TITLE-ABS-KEY (segmentation) OR TITLE-ABS--KEY (classification) OR TITLE-ABS-KEY ("data mining") OR TITLE-ABS-KEY ("data patterns")) AND (TITLE-ABS-KEY (omnichannel) OR TITLE-ABS-KEY (multichannel) OR TITLE-ABS-KEY (e-commerce) OR TITLE-ABS-KEY (sales)) AND (TITLE-ABS--KEY (b2b) OR TITLE-ABS-KEY (business-to-business) OR TITLE-ABS-KEY ("business to business"))

Source: own work.

Review execution

It was assumed that articles published between 2008 and 2018 would be analysed. It was established that the search was to be conducted in an automated manner. The next step in the study was to develop the following criteria for selecting scientific papers for detailed analysis:

- The selected term is searched for as a whole phrase or a combination of words occurring in it.
- The term must be included in the title or abstract of an article.
- Only publications in English were analysed.
- Articles containing the term searched for in the abstract.

The criteria for excluding publications for in-depth analysis were as follows:

- Articles which proposed solutions for other areas (e.g. life sciences).
- Incomplete articles, such as abstracts only and extended abstracts.
- Articles with fewer than four pages, published as the so-called ShortPaper.
- After the search, ambiguous articles and/or those irrelevant for the ongoing study were also excluded.

After executing a search for the defined strings in the selected databases (Table 1), an initial set of nine papers was obtained (Table 2).

Table 2. List of publications following the study

No.	Author(s) (year)	Journal	Title	Method
1	Hosseini, S.M.S., Maleki, A., Gholamian, M.R. (2010)	Expert Systems with Applications	Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty	Case study
2	Webster, M., Cullen, A.J. (2007)	International Journal of Operations and Production Management	A model of B2B e-commerce, based on connectivity and purpose	Model
3	Simkin, L. (2008)	Journal of Business and Industrial Marketing	Achieving market segmentation from B2B sectorisation	Theory
4	Müller, J.M., Pommeranz, B., Weisser, J., Voigt, K. I. (2018)	Industrial Marketing Management	Digital, Social Media, and Mobile Marketing in industrial buying: Still in need of customer segmentation? Empirical evidence from Poland and Germany	Case study
5	Bohanec, M., Robnik-Šikonja, M., Borštnar, M.K. (2017)	Industrial Management and Data Systems	Decision-making framework with double-loop learning through interpretable black-box machine learning models	Model
6	Hosseini, M., Shabani, M. (2015)	Journal of Marketing Analytics	New approach to customer segmentation based on changes in customer value	Model
7	Thomas, R.J. (2016)	Journal of Business and Industrial Marketing	Multistage market segmentation: an exploration of B2B segment alignment	Case study
8	Sánchez-Hernández, G., Chiclana, F., Agell, N., Aguado, J.C. (2013)	Knowledge-Based Systems	Ranking and selection of unsupervised learning marketing segmentation	Theory
9	Exenberger, E., Bucko, J. (2020)	Agris On-line Papers in Economics and Informatics	Analysis of Online Consumer Behavior – Design of CRISP-DM Process Model	Model

Source: own work.

Review reporting

After analysing all the selected papers, the results were compiled to provide answers to the research questions. The next section presents a summary of the selected studies in terms of their defining characteristics and methodology adopted as well as discussing the main issues that emerged from the review. Then the promising research directions were highlighted with respect to the clustering and segmentation of customers and products in omnichannel retail companies.

2.3. Conclusions from the literature study

The completed systematic literature review study yielded results that answered the research questions posed. The machine learning techniques most commonly used for customer and product clustering and segmentation are as follows (RQ 1.1.):

- *k*-Means Clustering Algorithm (Müller, Pommeranz, Weisser, & Voigt, 2018; Webster & Cullen, 2007);
- Recency, Frequency, Money (RFM) analysis (Müller et al., 2018);
- Classification And Regression Tree (CART)(Webster & Cullen, 2007);
- Apriori Association Rule Mining (ARM) (Apriori) shopping basket analysis (Webster & Cullen, 2007);
- unsupervised learning methods to automatically generate a set of classifications (Exenberger & Bucko, 2020).

The analysis performed indicates CRISP-Data Mining as a comprehensive method providing a structured approach to data mining project planning (Exenberger & Bucko, 2020; Hosseini, Maleki, & Gholamian, 2010).

The analysis of the publications on the basis of the adopted research procedure allowed for the identification of key research directions related to clustering and segmentation. Most research (Bohanec, Robnik-Šikonja, & Borštnar, 2017; Hosseini & Shabani, 2015; Sánchez-Hernández, Chiclana, Agell, & Aguado, 2013; Webster & Cullen, 2007) (is currently focused on combining machine-learning techniques for the methods developed, or comprehensive frameworks using CRISP-DM (RQ 1.2). Segmenting customers based on multiple criteria may ultimately increase the effectiveness of association rules and thus enable the achievement of business objectives (Webster & Cullen, 2007). One important aspect of building frameworks is combining interpretable machine learning models with human understanding of the decision problem (RQ 1.2), demonstrating how to integrate interpretable knowledge-discovery technology recommendations into complex decision-making and continuously support organisational learning.

3. Case study research – stage two of the study

3.1. Research methodology

The rigour of conducting case study research has received little attention in the literature. The following is one of the case study research procedures according to R.K. Yin *Designing Case Studies* (2003). Conducting a study using this method consists of the following main stages: defining and designing the study, preparing for data collection, data collection and analysis, and analysis and inference (see Figure 2).

At the research design stage, attention should be paid to issues concerning the following aspects:

- formulating research questions (such as how and why),

- related statements, hypotheses (if any; not necessarily for exploratory case studies),
- logical connections between data and statements and hypotheses (e.g. pattern matching technique).

Preparation for conducting a case study involves:

- the researcher's skills (the ability to ask 'good' questions and interpret the answers is a prerequisite);
- the ability to listen and avoid falling into their own methodological traps or concepts;
- the ability to assimilate a great deal of information in an unbiased manner;
- adaptability and flexibility (new situations seen as opportunities rather than threats);
- without a good knowledge of the issues being researched, there is a danger of passing over important issues and not being able to determine whether a deviation is acceptable or even desirable;
- impartiality towards preconceived conclusions, including those drawn from theory – the researcher should be sensitive and prepared to respond to conflicting pieces of evidence;
- training and preparation for conducting a specific case study;
- development of a case study protocol.

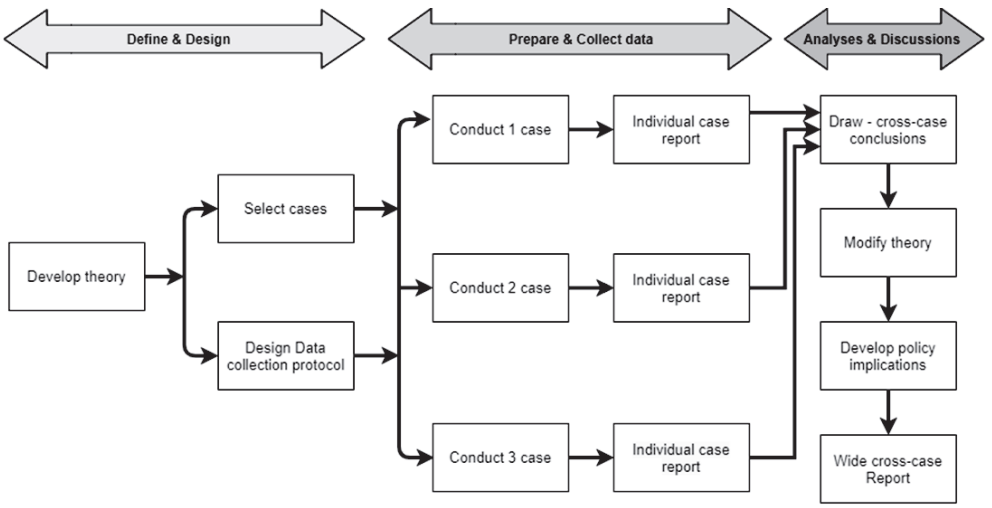

Fig. 2. Case study research: design and procedure

Source: (Yin, 2003).

Understanding and offering a consistent interpretation of usually disparate data sources (both qualitative and quantitative) is not easy. Repeatedly reviewing and

sorting extensive and very detailed data is an integral part of the analysis process. In joint case studies, it is useful to first analyse the data relating to each component case and only then make comparisons between the cases. Attention should be paid to the differences in each case and, where appropriate, the relationship between the different causes and effects (Miles, 1994). Data will need to be organised and coded to allow key issues, both from the literature and emerging from the dataset, to be easily found at a later stage. The final part of the report contains the conclusions, which should present a comprehensive view of the exploration of the main theme of the study carried out and the achievement of its objective.

3.2. Execution of the study and results of stage two – case study research

The study was carried out in two B2B companies with omnichannel sales. The companies examined had varying levels of maturity in terms of their use of IT solutions. Selecting companies with different levels of maturity was a deliberate measure aimed to gain a broader perspective on the research questions posed.

The study used a technique consisting in interviews with company employees at different levels of the organisational structure – from senior management to specialists. All the participants in the survey were asked the same questions, the most important of which included the following:

- What business decisions are made based on data analysis?
- Is the company involved in the area of product or customer segmentation?
- How does the company assess the use of data analysis tools related to segmentation and clustering of products or customers?
- In which area does the company have the greatest business potential regarding segmentation and clustering of products and customers?

In the course of the survey, six people from two companies were interviewed. The conclusions from the study are presented in the section below.

3.3. Conclusions from the case study research

The completed case study research yielded results that answered the research questions posed. In the companies examined, the main issues related to clustering and customer segmentation are customer service and matching their product offer to the customer. They base their strategies on the development of perfect customer service whose quality is constantly being improved and presented as part of their market offer. The company's regular and potential customers present in the market view this offer from the viewpoint of its potential value to their individual goals. Segmentation and clustering enable easier strategy execution by building customer segments with similar characteristics or purchase histories.

In practice, one can also see that managers are changing the structures of the studied companies so that they become customer-centric. Customer-focused processes and skills are also being developed. The management of the sales teams' activities is carried out with a focus on individual customer segments and groups,

which are treated specifically, and the most appropriate mix of services and products is selected to enable gaining and retaining customers.

The surveyed companies use simple customer segmentation based on RFM analysis (standing for Recency, Frequency, Monetary). RFM segmentation is based on three variables: date of last customer purchase (Recency), frequency of purchases (Frequency), value of all transactions made (Montetary).

The problem of classifying two user sessions in an online store, a shopping session and a browsing session was analysed with the K-nearest neighbour method (using Euclidean distance) as the most effective in terms of shopping and overall predictions.

Market Basket Analysis is a technique used by survey participants to discover links between products. It works by looking for combinations of products that occur together frequently in transactions, namely it allows retailers to identify relations between the items that people buy. Association Rules are widely used to analyse retail basket or transaction data, and are intended to identify strong rules discovered in transaction data using measures of interestingness, based on the concept of strong rules.

An area that was indicated as strongly developing and expected by the surveyed companies is the more accurate segmentation and clustering of customers and products. Currently, the number of segments in a dimension or dimensions is small due to the need for manual analysis and recommendations. The expected solution is to use machine learning to build micro-segments to better categorise customers or products. The segments will contain a small number of customers or products compared to the classic methods, which will result in their micro scale, i.e. being micro-segments.

4. Conclusions and future works

The article presents a study conducted to identify the clustering and segmentation needs in trade companies operating in the B2B sector and the machine-learning techniques currently used for clustering and segmentation. The study was divided into two stages. In the first stage, using the systematic literature review technique, an analysis of machine learning techniques and research directions related to clustering and segmentation was conducted. As a result, the following machine--learning techniques for segmentation and clustering were identified: the K-Means Clustering Algorithm (Müller et al., 2018; Webster & Cullen, 2007); RFM (Recency, Frequency, Money) analysis (Müller et al., 2018); Classification And Regression Trees (CART) (Webster & Cullen, 2007); Apriori Association Rule Mining (ARM) (Apriori) shopping basket analysis (Webster & Cullen, 2007); and unsupervised learning methods to automatically generate a set of classifications (Exenberger & Bucko, 2020). Major research directions in the literature include combining interpretable machine-learning models with human understanding of the decision

problem, and outlining how to integrate interpretable knowledge-discovery technology recommendations into a complex decision-making process and continuously support organisational learning.

In the second stage, in a preliminary study of two multichannel sales companies operating in the B2B segment, the results show that customer/product segmentation and clustering is used to improve customer service levels, which is reflected in well-tailored offers for customers. A research area indicated by the business entities under analysis was identified, namely the more accurate segmentation and clustering of customers and products. From the enterprise's perspective, such a solution should, based on advanced data analytics models, create micro-segments that allow for assigning the customer to one of them, as well as enabling better service for the customer and product offer better tailored from the enterprise's perspective.

An area that was indicated as strongly developing and expected by the surveyed companies is the more accurate segmentation and clustering of customers and products. Currently, the number of segments in a dimension/s is small due to the need for manual analysis and recommendations. The expected solution is to use machine learning to build micro-segments to better categorise customers or products. The segments will contain a small number of customers or products compared to the classic methods, which will result in their micro scale, i.e. being micro-segments.

Further work will focus on building a method to enable the micro-segmentation of customers and products, and an algorithm to transfer customers between micro--segments in the sales process in order to maximise their value to the company.

References

Bąska, M., Dudycz, H., & Pondel, M. (2019). Identification of advanced data analysis in marketing: A systematic literature review. *Journal of Economics & Management*, *35*, 18-39.

Bohanec, M., Robnik-Šikonja, M., & Borštnar, M. K. (2017). Decision-making framework with double-loop learning through interpretable black-box machine learning models. *Industrial Management & Data Systems*, (7).

Denyer, D., & Tranfield, D. (2009). Producing a systematic review. In D. Buchanan, & A. Bryman (Eds), *The Sage handbook of organizational research methods* (pp. 671-689). London: Sage Publications Ltd.

Exenberger, E., & Bucko, J. (2020). Analysis of Online Consumer Behavior-Design of the CRISP-DM Process Model. *Agris On-Line Papers in Economics & Informatics*, (3).

Hofmann, E., & Bosshard, J. (2017). Supply chain management and activity-based costing: Current status and directions for the future. *International Journal of Physical Distribution and Logistics Management*, *47*(8), 712-735.

Hohenstein, N. O., Feisel, E., Hartmann, E., & Giunipero, L. (2015). Research on the phenomenon of supply chain resilience: a systematic review and paths for further investigation. *International Journal of Physical Distribution & Logistics Management*, *45*(1-2), 90-117.

Hosseini, S. M. S., Maleki, A., & Gholamian, M. R. (2010). Cluster analysis using the data mining approach to develop CRM methodology to assess customer loyalty. *Expert Systems with Applications*, *37*(7), 5259-5264.

Hosseini, M., & Shabani, M. (2015). A new approach to customer segmentation based on changes in customer value. *Journal of Marketing Analytics*, *3*(3), 110-121.

Kitchenham, B. (2004). *Procedures for performing systematic reviews* (Joint Technical Report TR/SE-0401 and NICTA 0400011T.1 2004-07), Keele University.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. London: Sage Publications Ltd.

Müller, J. M., Pommeranz, B., Weisser, J., & Voigt, K. I. (2018). Digital, social media, and mobile marketing in industrial buying: Still in need of customer segmentation? Empirical evidence from Poland and Germany. *Industrial Marketing Management*, *73*, 70-83.

Sánchez-Hernández, G., Chiclana, F., Agell, N., & Aguado, J. C. (2013). Ranking and selection of unsupervised learning marketing segmentation. *Knowledge-Based Systems*, *44*, 20-33.

Simkin, L. (2008). Achieving market segmentation from B2B sectorisation. *Journal of Business & Industrial Marketing, 23*(7).

Thomas, R. J. (2016). Multistage market segmentation: an exploration of B2B segment alignment. *Journal of Business & Industrial Marketing*, *31*(7).

Unity Group. (2021). *Raport: Polski rynek B2B vs. Cyfrowy Megatrend. Sytuacja branży post COVID-19*. Retrieved April 4, 2021 from https://www.unitygroup.com/wp-content/uploads/2021/03/Polski-rynek-B2B-vs.-Cyfrowy-Megatrend-Sytuacja-branzy-post-COVID-19.pdf

Webster, M., & Cullen, A. J. (2007). A model of B2B e-commerce, based on connectivity and purpose. *International Journal of Operations & Production Management, 27*(2).

Yin, R. K. (2003). Designing Case Studies. Qualitative Research Methods, 359-386. Retrieved from https://books.google.pl/books?id=KdneDQAAQBAJ&lpg=PA359&ots=gOPkErCfgW&dq=4.%09Yin%2C%20R.%20K.%20(2003).%20Designing%20Case%20Studies.%20Qualitative%20Research%20Methods%2C%20359-386.&lr&hl=pl&pg=PR4#v=onepage&q&f=false