Applied Soft Computing Journal 182 (2025) 113559

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Multi-criteria selection of data clustering methods for e-commerce personalization

Elżbieta Pawełek-Lubera [c], Mateusz Przyborowski [b], Dominik Ślęzak [b,d], Adam Wasilewski [a,c,*]

[a] Faculty of Management, Wrocław University of Science and Technology, Wyb. Wyspiańskiego 27, Wrocław, 50-370, Poland
[b] Faculty of Mathematics, Informatics, and Mechanics, University of Warsaw, Stefana Banacha 2, Warszawa, 02-097, Poland
[c] Fast White Cat S.A., Fabryczna 6, Wrocław, 53-609, Poland
[d] QED Software Sp. z o.o., Mazowiecka 11/49, Warszawa, 00-052, Poland

ARTICLE INFO

Keywords:
Multi-criteria decision methods
Data clustering
User interface
E-commerce
PROMETHEE II
Approximated Gaussian Mixture Model

ABSTRACT

E-commerce platforms increasingly rely on personalization to improve the user experience and drive sales, requiring efficient data clustering methods to segment users based on their behavior and preferences. However, due to the many data clustering techniques available, the key decision problem is to choose the optimal grouping method. Decision making based on single decision factors, although widely used, can lead to wrong decisions, so it is worth considering multi-criteria analysis tailored to the specifics of e-commerce customer clustering. Through extensive experiments on real e-commerce datasets, the study demonstrates the strengths and limitations of selected data clustering techniques (including the Approximated Gaussian Mixture Model, which was found to be superior to the classical Gaussian Mixture Model), considering different decision criteria related to various aspects of quality. The results provide valuable insights for e-commerce practitioners seeking to optimize their personalization strategies and ultimately suggest that a novel adaptation of the PROMETHEE II method can provide a robust framework for making informed decisions about selection of data clustering algorithms.

1. Introduction

E-commerce customer behavior and preferences are critical for companies seeking to improve customer satisfaction, increase sales, and maintain a competitive edge. However, data analysis and inference are facing new challenges, primarily related to the amount of information collected and processed. One example is customer segmentation, which in traditional solutions is based on the Recency, Frequency, and Monetary (RFM) approach. Such an approach primarily considers data on customers' final purchase decisions. However, when it is necessary to consider the entire user activity in the online store, taking into account each step of the customer journey in detail, the amount of data collected rapidly increases, which affects the ability to process it. The solution can be to use machine learning or soft computing methods, for example, to segment e-commerce customers, analyze their growth potential, improve marketing efforts and identify new markets [1]. In general, such techniques can be characterized as aimed at dividing customers into distinct groups based on various attributes such as browsing behavior, purchase history, demographics, localization, and psychological aspects.

Among potential applications of user grouping in online stores are personalization activities. The most popular are recommendation systems (e.g., of products or services). However, there are also more sophisticated applications allowing for customization not only of the content presented but also of the layout. Personalized User Interfaces (PUIs) have emerged as a critical solution for enhancing the user experience by tailoring the online shopping environment to individual needs and preferences. The cornerstone of developing effective PUIs is robust customer segmentation, which enables companies to understand and anticipate the unique needs of their user base. Initial research on using data clustering to prepare groups of customers to be served a dedicated UI variant has shown the great potential [2]. The most promising – according to those previous studies – data clustering methods (K-means and Gaussian Mixture Model, further abbreviated as GMM) were selected for the analysis described in this paper. Their modified versions (like Fuzzy K-means) were analyzed as well. The Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH) method, which is a hierarchical data clustering approach, was chosen as an additional benchmark. Finally, an approach previously used in other application areas, which is a modification of the GMM algorithm – Approximated Gaussian Mixture Model, further abbreviated as (GMM+) [3] – was adopted for this study. This method, similarly to BIRCH, is of particular

* Corresponding author at: Faculty of Management, Wrocław University of Science and Technology, Wyb. Wyspiańskiego 27, Wrocław, 50-370, Poland.
E-mail address: adam.wasilewski@pwr.edu.pl (A. Wasilewski).

https://doi.org/10.1016/j.asoc.2025.113559
Received 12 August 2024; Received in revised form 22 May 2025; Accepted 28 June 2025
Available online 16 July 2025

interest because of the speed of calculations, which is crucial for the application field considered in this paper.

The paper includes a novel approach to multi-criteria evaluation of data clustering methods aimed at developing dedicated user interfaces (UIs) in e-commerce, along with its use to comprehensively compare the features of the authors' GMM+ model with some other popular clustering algorithms. Recognizing that no single data clustering algorithm is a one-size-fits-all solution for all applications and data types, a comprehensive evaluation framework that considers multiple decision criteria is presented. Traditional evaluation metrics often emphasize statistical measures of cluster quality, such as consistency and separation, which may not fully capture the practical implications for UI personalization. In addition, considering single-criteria analysis may lead to erroneous decisions, negatively affecting the business efficiency of multi-variant UI systems. Therefore, the proposed approach includes a broader set of criteria covering both technical performance and business relevance and is based on one of the Multi-Criteria Decision Methods (MCDMs) – Preference Ranking Organization Method for Enrichment Evaluation (PROMETHEE) II. This is an approach that relies on pairwise comparisons of alternatives for each decision criterion, which, if used to select the optimal data clustering method for a multivariate e-commerce UI, would complicate the computational process and make the results difficult to interpret. For this reason, we propose to adapt the classical solution and tailor it to the specifics of the data clustering algorithm comparison.

The following research questions were formulated to conduct this study from the perspective of the proposed adapted version of PROMETHEE II:

RQ.1 How can recommendation of the best algorithm for clustering e-commerce behavioral data for serving dedicated UI variants depend on the assumed number of the resulting clusters?

RQ.2 Are the recommendations for optimal methods of grouping e-commerce customers obtained after using the adapted PROMETHEE II method confirmed by the knowledge of the specifics of the analyzed algorithms?

RQ.3 Does the choice of weights of decision factors affect recommendations?

The contribution of our research is threefold. First, a novel multi-criteria framework for evaluating data clustering algorithms, specifically for e-commerce personalization, is introduced. Its originality lies in the holistic integration of not only technical performance and standard clustering quality indicators but also resource usage constraints and a set of newly defined business context metrics tailored to the practical needs of developing PUIs. It represents an advancement over existing evaluation approaches and integrates unique quality indicators and allows an accurate evaluation from both technical and business perspectives. It is worth noting that the analysis was conducted on the basis of detailed behavioral data covering all e-commerce customer activities in detail, which further distinguishes the research described from analyses based only on the final results (e.g. orders placed).

As the second contribution, we propose a brand new adaptation of the PROMETHEE II method. This adaptation is novel in its two-step approach, which allows for the systematic comparison and selection of clustering options (algorithm and number of clusters) even when the final number of clusters is not initially defined by business requirements – a common challenge in this domain. It extends typical applications of PROMETHEE II in algorithm selection. This adaptation provides a significant methodological advancement for business cases where the number of clusters is not clearly defined.

As the third contribution, the proposed framework is applied to provide the first comprehensive, multi-criteria comparison of the authors' previously introduced GMM+ model against other leading clustering algorithms on detailed real-world e-commerce behavioral data. The analysis contributes uniquely to the field by offering the rigorous validation of GMM+, providing practical insights that go beyond theoretical algorithm comparisons.

The experimental results obtained on the real-world data confirmed the usefulness of the proposed approach for comparing clustering algorithms and justified the importance of considering the business context as a decision criterion. In addition, the study showed that the GMM+ approach, previously introduced by the authors, locates high in the rankings obtained after applying the modified PROMETHEE II method. In particular, it is better than the standard GMM algorithm and its standard extensions/accelerators.

2. Literature review

2.1. Personalization of e-commerce user interfaces

Many studies have addressed issues related to the personalized user interfaces of web-based systems, emphasizing aspects such as human-centeredness, satisfaction, and loyalty in the context of ensuring privacy and security [4]. Studies of literature show that activities can be architectural, relational, instrumental, or commercial. However, the latter two are the most popular in academic literature [5]. With the growth of e-commerce, it is becoming increasingly important to understand the factors that contribute to successful personalization, which will positively impact the customer experience and the business efficiency of platform owners. Other studies indicate that website design significantly affects customer satisfaction and loyalty, and usability plays a significant role in the shopping process [6]. A well-designed interface not only facilitates navigation but also improves the overall shopping experience, increasing customer satisfaction [7] and loyalty [8]. Other studies confirm this, showing that satisfaction and the intent to revisit a personalized website can be influenced by information, navigation, and presentation [9]. It is worth noting that personalization can apply to the entire layout and its elements, such as product carousels [10]. However, the main goal is always to tailor it to users' behaviors and needs.

Another key aspect of personalization in e-commerce is the use of recommendation systems. These systems analyze user behavior and preferences to provide tailored product suggestions, making it easier for customers to make purchasing decisions [11]. Studies show that personalized recommendations can significantly impact the effectiveness of an online store by helping users find products that meet their needs [12]. Additionally, the potential of machine learning techniques [13] and big data [14] for taking personalization to the next level is being recognized.

When analyzing personalization, it is also important to consider its ethical implications. Concerns about data privacy and security are significant to customers, who are increasingly aware of how their data is used by e-commerce platforms [15]. Transparency in data collection and use can increase user trust and lead to the development of long-term relationships between consumers and online retailers [16]. The perception that personalization does not violate privacy has a significant impact on customer perceptions, satisfaction, and loyalty [17].

In conclusion, the personalization of user interfaces in e-commerce is very important due to the need to continuously ensure customer satisfaction and loyalty. By implementing this approach, it is possible to improve the shopping experience for users. However, when analyzing personalization strategies, legal and ethical issues cannot be overlooked, especially given the dynamics of change in e-commerce.

2.2. Data clustering in e-commerce

The analysis of e-commerce user behavior requires considering specific problems. The biggest challenge is the number of customers who buy online. Worldwide there are more than 5 billion potential customers [18], each with their own characteristics, expectations and preferences. At the scale of a single online store, the numbers are much smaller, but still too large to consider analyzing each user's behavior individually. The solution is customer segmentation, a marketing technique that has been used for many years [19]. Originally, the dominant method of segmentation was the use of decision rules [20] to divide customers by demographics, geography, technology, and so on. Over time, increasing data collection and processing capabilities have made it possible to use machine learning techniques for this purpose, including data clustering algorithms, an example of unsupervised learning applications [21]. Today, this approach is widely used in e-commerce in a variety of ways [22]. Customer segmentation enables the personalization of communications, affecting both the content presented, offers, and the layout of messages [23].

Various data clustering algorithms are used to segment users on the base of e-commerce data. The most common approach is K-means [22], which is used to analyze selected customer characteristics (e.g. purchases and the number of visits [24]) and continuous (clickstream) data [25]. Other data clustering algorithms used for e-commerce segmentation include:

- The aforementioned BIRCH [26] and GMM [27].
- Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [28].
- K-medoids [29].
- Self-Organizing Maps (SOM) [30].
- Fuzzy clustering approaches [31,32], etc.

The main way to use clustering results refers to various types of recommendations, especially of products [33,34], reading content [35], and logistics [36]. Data clustering can also be the foundation for more complex personalization solutions that can serve customers with multivariant, behaviorally tailored UIs [37]. However, regardless of the purpose of clustering, the system design challenge is to select the right algorithm and its parameterization. As data clustering approaches vary, so do their effects. As a result, the performance of the entire system (e.g., product recommendations) depends on the initial decision about which datasets to use and how to group users.

2.3. GMM in data clustering

As one of our achievements in this paper refers to the application of the aforementioned GMM+ approach, let us focus first on GMM. It is a probabilistic model used in data clustering tasks, particularly when the distribution of data points is not known in advance. GMM assumes that the data is generated from a mixture of several Gaussian distributions, each representing a cluster in the data. GMM is a soft clustering algorithm, meaning that it assigns each data point a probability of belonging to each cluster rather than a hard assignment to a single cluster. This allows for more flexibility in capturing complex data patterns and dealing with overlapping clusters.

In data clustering, GMM can be used to assign data points to different clusters based on their probability of belonging to each cluster. This is done by iteratively updating the parameters of the Gaussian distributions (mean and covariance) to maximize the likelihood of the data given the observed model. Several examples of finite mixture models for cluster analysis are given in [38]. Such models are growing in popularity, largely due to the generality of the assumption about the probability distribution of the data.

In its most basic formulation, GMM is a relatively simple model that can be adapted. There are several methods to apply it to the problem as well as to overcome its shortcomings in the known literature:

- In [39], the authors propose a likelihood-based metric for evaluating GMM clustering.
- In [40], the authors try to distinguish between the number of Gaussian components and the number of clusters using the entropy criterion.
- In [41], the authors describe the use of GMMs as a variable selection tool for data clustering.

More sophisticated approaches are also popular, such as [42] with strict application to Variational Autoencoders, i.e., neural networks with a GMM used as a prior distribution.

2.4. Selection of data clustering algorithms

Attempts to address this problem of choosing the optimal solution can be found in publications comparing different data clustering methods. Literature examples of such analyses that are based on e-commerce datasets are shown in Table 1.

A characteristic feature of studies comparing data clustering methods, especially in the area of e-commerce, is the reduction of decision factors, in extreme cases to a single indicator of the data clustering quality (such as Davies Bouldin Score, Silhouette Score). However, the selection of the optimal solution requires the analysis of many more factors in order to evaluate the different aspects of the methods to be compared [48]. In such a situation, a worthwhile option to consider is the use of multi-criteria decision methods (MCDM), which allow for objectively considering many features of available data clustering algorithms.

2.5. Applications of MCDM

MCDM is a set of methods and processes used to evaluate and prioritize multiple competing criteria in decision making. They are applied to support decisions in the presence of multiple, usually conflicting, factors and characteristics. Among the most important MCDM concepts are [49]:

- Alternatives — options or decisions available to the decision maker.
- Criteria — parameters (characteristics, indicators, metrics, etc.) for evaluating alternatives.
- Weights — the importance (objective or subjective) of each criterion to the decision maker.
- Decision Matrix — a summary with a ranking of alternatives in relation to the criteria.

There are several solutions classified as *classical MCDM approaches* [49], including:

- *Elimination and Choice Expressing Reality* (ELECTRE).
- The aforementioned PROMETHEE approach.
- *Multi-Attribute Utility Theory* (MAUT).
- *UTilités Additives* (UTA).
- *Measuring Attractiveness by a Categorical Based Evaluation Technique* (MACBETH).

It is also worth mentioning other methods that are often used in practice, such as *Technique for Order of Preference by Similarity to Ideal Solution* (TOPSIS) [50], *Weighted Sum Model* (WSM) [51], *Data Envelopment Analysis* (DEA) [52], *Complex Proportional Assessment* (COPRAS) [53], and *Vlsekriterijumska Optimizacija I Kompromisno Resenje* (VIKOR) [54].

These techniques are particularly valuable in complex decision scenarios where multiple factors must be considered simultaneously [55]. This approach is applied in many fields such as environment [56], energy [57,58], materials [59,60], engineering [61], medicine [62], information and communication technologies [63], economics [64,65],

Table 1
Examples of comparative studies of data clustering algorithms in e-commerce.

Compared algorithms	Dataset(s)	Conclusions	Reference
K-means, GPHC	Feedback questionnaires from customers	On unbalanced and high dimensional scenarios, GPHC effectively performs the data clustering task	[43]
K-means, Agglomerative, BIRCH, Spectral	Products, personal information, customer decisions, geographic factors	K-means method had the smallest sum of squares of errors	[44]
K-means, Hierarchical Clustering	Geographic, psychological, behavioral, and demographic factors	K-Means gives most accurate results	[45]
GMM, K-Means, BIRCH	Personal data, introduced spending score	K-Means clustering was the best data clustering algorithm in terms of Davies Bouldin Score	[46]
K-means, GMM, DBSCAN, Agglomerative, BIRCH	UK-based online retail dataset	GMM outperformed other approaches in terms of Silhouette Score	[47]

etc. However, although multi-criteria analysis could provide the basis for selecting the optimal data clustering method, only several publications focus on such solutions. Some interesting applications of MCDM in data clustering include:

- Application and comparison of six data clustering algorithms using three MCDM methods (TOPSIS, DEA, VIKOR) in financial risk analysis [66].
- Comparison of six data clustering algorithms using three MCDM methods (TOPSIS, COPRAS, WSM) on 10 different datasets from the University of California Machine Learning Repository [67].
- Using the MCD (PROMETHEE) method to solve the cluster ordering problem (e.g., from best to worst) [68].
- The use of multi-criteria analysis in the process of data clustering. In this case, MCDM is not used to evaluate the clustering results, but to support the computations [69].

In addition, approaches based on "fuzziness" in multi-criteria analysis are gaining popularity, as illustrated by various practical applications [70–72].

It is worth noting that among the examples of using MCDM to develop or evaluate data clustering methods, there are no applications directly related to customer behavior in online stores. However, this does not mean that multi-criteria analysis is not used in e-commerce. For example, it can be an important complement to the recommendation systems [73] that are an integral part of many e-commerce platforms. Sales [74] and logistics [75] in e-commerce can also be implementation areas for this approach.

In summary, it can be concluded that there is a lack of comprehensive studies that analyze data clustering algorithms using MCDM, located in the business context of e-commerce customer behavior. This is the research gap that this paper directly addresses by proposing and validating a novel, comprehensive MCDM-based methodology tailored for e-commerce personalization. Our work not only continues and formalizes initial analyses [76] but critically introduces a unique adaptation of PROMETHEE II and applies it to evaluate a set of algorithms, including the aforementioned recent GMM+ modification of the standard GMM model [3], offering a new, structured approach to a complex decision problem in the e-commerce domain.

3. Method

This section outlines the systematic methodology employed to identify and evaluate optimal data clustering methods for personalizing e-commerce user interfaces. The proposed approach, illustrated in Fig. 1, unfolds in several key stages designed to provide a robust and contextually relevant selection framework. The first stage allows for the gathering and anonymizing extensive behavioral data from a live e-commerce platform. This raw data then underwent a thorough preprocessing phase to create a suitable feature space for user segmentation. Following this, a diverse set of established and modified data clustering algorithms, including K-means, Fuzzy K-means, GMM, its aforementioned modification GMM+, and BIRCH, was applied to the data, varying the number of target clusters to explore different segmentation granularities. For additional reference, we also examined one more technique for accelerating/approximating the standard GMM (the details can be found in further sections), but we eventually decided to exclude it because of poor quality results.

A central and novel component of our methodology is the adaptation of the PROMETHEE II multi-criteria decision making technique. This adaptation was specifically designed to address the challenge of selecting the best combination of clustering algorithm and number of clusters, particularly when the ideal number of segments is not known a priori. This approach allowed us to evaluate the clustering options against a comprehensive suite of criteria encompassing context-free quality metrics, resource usage, and business-context relevance. The final stage involved a sensitivity analysis to examine the stability of recommendations under different criteria weighting schemes. The subsequent subsections provide detailed descriptions of the data and data clustering procedures, as well as the multi-criteria analysis framework, including the adapted PROMETHEE II method and the specific decision criteria utilized.

3.1. E-commerce customer behavioral data clustering

3.1.1. The data

The data is collected based on activity in an online store offering children's clothing. Each person who has an account on the site and logs in during an activity session consents to the collection of certain data, which is then aggregated and stored on an external server for sales analysis and predictive modeling. This source served as the data provider for the experiments.

The dataset used for data clustering consists of 4 months of user activity. 818,152 different customers were recorded, over 57% of users are responsible for a single visit, and over 90% of users made less than 6 visits. Only 2% of users visited the site at least 15 times. The total number of visits is 2,301,303, with an average of 19,177 visits per day. Detailed logs include information about the device used (type, resolution, browser, etc.), country and language, or previous visits. Each visit is described with details that include the type of pages loaded during the user's activity, taking into account features such as the category of object viewed, the time spent, or whether the activity ended with a purchase. Another important feature is derived from the items viewed or purchased, including their categories and characteristics that can be compared to other items, such as color or gender. Products noted during user activity are matched with existing historical products, allowing for better segmentation of users at a specific product category level.

The pre-processing of the data consists of aggregation so that lists of sessions are translated into a list of activities for each user. Each such list is then summarized with the aforementioned statistics, which differ depending on the type of feature. Finally, appropriate transformations and scaling are applied.

Fig. 2 shows an example of the partial user representation used during the processing of the dataset. The summaries differ in the number of visits, so they are represented by vectors.

Fig. 1. General flowchart of the approach.

	lastActionTimestamp	firstActionTimestamp	visitCount	totalEcommerceRevenue	actionType
00002e39-057c-4422-9bdb-566cbb7e49da	[1704275079, 1704473899, 1704396592]	[1704274242, 1704473899, 1704396525]	[1, 3, 2]	[85.71, 85.71, 85.71]	[{'action': 54, 'ecommerceOrder': 1, 'event': ...
0000392a-73cf-4f06-a31a-660d46de7003	[1709376895]	[1709375366]	[1]	[0]	[{'action': 106, 'ecommerceAbandonedCart': 1, ...
000071f1-4165-48e0-9ab2-0b2bbb07838a	[1711350530]	[1711350503]	[2]	[0]	[{'action': 8, 'event': 2}]
00007b8d-c036-4845-bda0-35e149753636	[1711469522, 1712006458]	[1711468808, 1712005502]	[1, 2]	[0, 0]	[{'action': 45, 'event': 10}, {'action': 82}]
00008021-993e-4474-af55-1b7ee56440f0	[1704999542]	[1704999347]	[1]	[0]	[{'action': 21, 'event': 10}]
000094d8-9674-47d3-a0a8-eeddd92b81f5	[1713955965, 1713943419]	[1713955964, 1713943111]	[2, 1]	[0, 0]	[{'action': 5}, {'action': 51, 'event': 1}]
0000d572-eac4-44d3-a4bf-61ef4b4224a8	[1709019473]	[1709019379]	[1]	[0]	[{'action': 10, 'event': 5}]
0000d9cf-6296-4cc2-b50f-150e7b03577d	[1710071141, 1710102010, 1710168842, 1709588796]	[1710071140, 1710101390, 1710168839, 1709588752]	[2, 3, 4, 1]	[0, 0, 0, 0]	[{'action': 2, 'event': 1}, {'action': 4, 'eve...

Fig. 2. Example of statistics (rows) per user activity (columns).

3.1.2. The algorithms

Data clustering was performed using 5 algorithms:

- **K-means** [77] is an iterative data clustering algorithm that assigns each data point to the cluster with the closest mean value.
  1. Initialize cluster centroids.
  2. Assign each data point to the nearest centroid based on a Euclidean distance metric.
  3. Update the centroids by calculating the mean of all data points assigned to each cluster.
  4. Repeat steps 2 and 3 until convergence is reached.

  K-means effectively minimizes intra-cluster sum of squares:

$$\arg\min_{C} \sum_{i=1}^{k} |C_i| Var\, C_i \quad (1)$$

  i.e., it searches for the clustering $C = \{C_1, \ldots, C_k\}$ that achieves the minimum, where C_i is a set of data points belonging to the ith cluster.
- **Fuzzy K-means** [78,79] is an extension of K-means clustering that assigns each data point a vector of membership values for each cluster. The key step is to update centroids by computing the weighted mean of all data points, where the weight w_{ij} indicates the degree to which a point x_i comes from a cluster C_j. The algorithm effectively minimizes a weighted intra-cluster sum of

squares:

$$\arg\min_{C} \sum_{i=1}^{k} \sum_{j=1}^{n_i} w_{ij}^{m} \|x_j - c_i\|^2 \tag{2}$$

With c_i being the center of a cluster C_i, n_i being size of the ith cluster and weights:

$$w_{ij} = \left(\sum_{p=1}^{k} \frac{\|x_j - c_i\|^2}{\|x_j - c_p\|^2}\right)^{\frac{2}{1-m}} \tag{3}$$

The parameter $m > 1$ is responsible for a fuzzy part of the algorithm, where $m \to 1$ converges to the default K-means and $m \to \infty$ results in increasingly fuzzy clusters.

- **Gaussian Mixture Model (GMM)** [80] is a probabilistic model that assumes that each data point follows a Gaussian distribution. The algorithm estimates the parameters of these distributions using the Expectation–Maximization (EM) algorithm.

  1. Initialize the distribution and weight parameters.
  2. E-step: calculate, based on the current parameters, the probability that each data point belongs to each component.
  3. M-step: update the parameters based on the E-step.
  4. Repeat steps 2 and 3 until convergence is reached.

  GMM assigns probabilities to the data points that belong to each cluster, rather than hard assignments as in K-means. This provides a more nuanced understanding of the data and can be useful in cases where data points clearly do not belong to a single cluster. Furthermore, unlike K-means, which assumes spherical clusters of similar size, GMM allows for clusters with different variances and sizes by estimating additional parameters for these two characteristics.
- **The Mini-Batch Gaussian Mixture Model (Mini-Batch GMM)** [81,82] is a scalable modification of GMM that allows it to handle large datasets more efficiently. It modifies the standard EM algorithm used in GMM by operating on small random subsets (mini-batches) of the data rather than the full dataset at each iteration. Instead of computing expectations (E-step) and parameter updates (M-step) over the entire dataset, Mini-Batch GMM uses a small batch of samples. This reduces memory usage and computation time, making it suitable for streaming data or large datasets. Unlike full EM which guarantees convergence, Mini-Batch GMM trades off convergence guarantees for faster, approximate solutions.
- **Approximated Gaussian Mixture Model (GMM+)** [3] is a modification of GMM proposed in earlier work that requires less computation for EM convergence. While GMM+ was demonstrating advantages on open datasets, a key novel contribution of the current study is its first comprehensive, multi-criteria comparative evaluation against a suite of established data clustering algorithms using real-world, complex e-commerce behavioral data. The key observation is that most of the data points close to the mean of a given component in the first few steps remain in the same cluster after all EM steps, but in GMM all these points are used to compute estimates, again and again. In GMM+, these points are merged into a single "granule" that behaves like a single data point with a weight equal to its cardinality. Given d-dimensional data point x, mean μ and covariance matrix Σ, with probability p we have

  $$(x-\mu)^T \Sigma^{-1} (x-\mu) \le \chi_d^2(p) \tag{4}$$

  where $\chi_d^2(p)$ is the quantile function of the chi-squared distribution with d degrees of freedom. This allows us to bound the number of data points that can be merged into the granule. It is worth noting that GMM+ not only speeds up the necessary computations but also acts as a regularization term, which is especially visible for datasets where multiple data points are very similar. In [3], the authors applied this concept to an open dataset to demonstrate the quality and computational speed of the proposed algorithm. The use of GMM+ on a real dataset and the conclusion of its efficiency is novel, both in terms of speeding up the computational phase of the algorithm and its effectiveness compared to the classical GMM method. To some extent, GMM+ can be also referred to the ideas of rough clustering [83], whereas the aforementioned granules correspond to the cluster lower approximations — the subsets of data points which are for sure in the given clusters.
- **BIRCH** [84] is a hierarchical clustering method that builds a tree structure to represent the data, with each node in the tree containing a cluster of data points. The algorithm represents the data using a set of sub-clusters, which are then hierarchically merged to form larger clusters based on their similarity.

3.2. Multi-criteria analysis of data clustering methods

3.2.1. PROMETHEE II

The choice of data clustering method can be considered in the context of the algorithm and its initialization parameters, in particular the number of resulting clusters (k). If the expected value of the parameter k in the decision process is known, the situation can be reduced to a multi-criteria analysis and a solution can be sought using selected techniques from the MCDM group. The situation is slightly different when searching for the optimal combination of the data clustering algorithm and the value of the parameter k, since treating each option as a decision alternative significantly increases the number of alternatives to be analyzed. Such a situation can be computationally inconvenient if the method involves comparing all alternatives in pairs (e.g., PROMETHEE, ELECTRE) and the number of decision criteria is large. To minimize this problem, an adaptation of a pairwise analysis approach tailored to the decision problems of selecting a data clustering algorithm and the number of clusters as the initial parameterization has been proposed. The PROMETHEE II technique was chosen for the study and embedded in the business context of choosing how to cluster e-commerce customers for serving dedicated user interface variants [37].

PROMETHEE II is designed to rank alternatives from best to worst based on multiple criteria and provides a complete ranking of alternatives, as opposed to PROMETHEE I, which provides a partial ranking. This approach is based on the concept of preference flows, which quantify the extent to which one alternative is preferred over another. By calculating the preference flow for each pair of alternatives, it determines the ranking of the alternatives based on their net preference flows [85].

In this approach, both alternatives (A) and criteria (C) can be expressed as an $n\times c$ decision matrix, which is evaluated in the following steps [86]:

1. Normalization of the decision matrix ((5) for beneficial criteria and (6) for non-beneficial criteria):

   $$R_{ij} = \frac{|X_{ij} - min(X_{ij})|}{|max(X_{ij}) - min(X_{ij})|} \tag{5}$$

   $$R_{ij} = \frac{|max(X_{ij}) - X_{ij}|}{|max(X_{ij}) - min(X_{ij})|} \tag{6}$$

   where X_{ij} represents the performance of the ith alternative with respect to the jth criterion; $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, c$;
2. For each criterion (j), calculation of the differences ($d_{ab|j}$) between each pair of alternatives (a, b):

   $$d_{ab|j} = R_{aj} - R_{bj} \tag{7}$$

   where $a, b \in A$, and $a \neq b$;

3. Calculation of preference function:

$$P_j(a,b) = \begin{cases} 0, & \text{if } d_{ab|j} \le 0 \\ d_{ab|j}, & \text{if } d_{ab|j} > 0 \end{cases} \tag{8}$$

4. Determination of the weights for the decision criteria (w_j) on the assumption that they are non-negative;
5. Calculation of the aggregated preference function:

$$\pi(a,b) = \frac{|\sum_{j=1}^{c} w_j \times P_j(a,b)|}{\sum_{j=1}^{c} w_j} \tag{9}$$

6. Calculation of the positive (ϕ^+) and negative (ϕ^-) outranking flows:

$$\phi^+(i) = \frac{1}{n-1} \sum_{\substack{b \in A \\ b \neq i}} \pi(i,b) \tag{10}$$

$$\phi^-(i) = \frac{1}{n-1} \sum_{\substack{b \in A \\ b \neq i}} \pi(b,i) \tag{11}$$

7. Calculation of the final outranking flow:

$$\phi(i) = \phi^+(i) - \phi^-(i) \tag{12}$$

8. Descending sorting of $\phi(i)$ values to rank alternatives.

In the case described (customer segmentation for a multi-variant user interface), the pairs (*data clustering algorithm*, *number of resulting clusters*) can be taken as the set of alternatives. The number of clustered e-commerce customer groups that are acceptable to the business will determine the number of options and computational complexity.

3.2.2. Adaptation of PROMETHEE II

The original and novel contribution of the paper is the proposed adaptation of the PROMETHEE II method to the specifics of evaluating data clustering methods, taking into account the specifics of data clustering algorithms and business needs, which can significantly improve the final recommendations. This novelty lies in its specific design to address two key use cases in evaluating data clustering methods for e-commerce personalization:

1. The expected number of resulting clusters is known and only the best data clustering algorithm needs to be identified.
2. The range of the number of clusters that may be useful for the business is known, and the optimal combination of the data clustering algorithm and the number of clusters in the range should be chosen.

According to the proposal, the first case is analyzed according to the standard procedure in PROMETHEE II, and the second is be processed in two steps:

1. Application of the PROMETHEE II method for all the number of clusters in the range that is acceptable for the business and selection of the best data clustering algorithm (alternatives are different data clustering algorithms, the number of clusters is fixed for each iteration).
2. The best algorithms (one for each number of clusters in the acceptable range) become alternatives, analyzed by the PROMETHEE II method, and from them the optimal combination of algorithm and number of clusters is selected.

Such a modified standard approach reduces the number of comparisons between alternatives, which decreases the computational complexity (especially with a large number of decision factors). An additional improvement is the ability to select the "best of the best" option when the number of outcome clusters is not predetermined.

3.2.3. Decision criteria

The second key element of the proposal, in addition to the identification of alternatives, is a set of decision criteria. It includes the commonly used data clustering quality metrics, the aspects of computational resources utilization, as well as the measures derived from business requirements and application-specific clustering effects. It was assumed that three measures would be selected in each group:

- Context-free metrics:
  - SS (Silhouette Score) – to be maximized.
  - DI (Dunn Index) — to be maximized.
  - DBI (Davies–Bouldin Index) — to be minimized.
- Resource-usage metrics:
  - AC (average usage of available processors in percent) – to be minimized.
  - MM (maximum memory usage in bytes) – to be minimized.
  - CT (clusterization time in seconds) – to be minimized.
- Business-context metrics:
  - CSI (Cluster Spread Index) – to be minimized.
  - AAD (Average Activity Diversity) – to be minimized.
  - SSC (Size of the Smallest Cluster) – to be minimized.

The indicators of the first group are popular ways to assess the quality of data clustering, but they do not take into account the context of using their effects. In addition to the three selected metrics, there are others (e.g. entropy, Caliński–Harabasz index), but they were discarded due to the fact that their calculation algorithm may favor the K-means method.

The second group of indicators are measures related to the use of computational resources. In this case, sample metrics related to different resources (memory, processor) and of different nature (average, maximum value) have been selected to emphasize the variety of options available. This list can be freely modified, as modern solutions allow comprehensive monitoring of server resources and the use of this data to assess the cost intensity of the data clustering algorithms analyzed.

The last group includes measures that address the business requirements of a specific cluster application. Since they represent groups of e-commerce customers to be served with dedicated UI variants, they should be characterized by:

- Minimal variation in the number of customers in the clusters, since the UI variants will then have a similar number of recipients.
- High distinguishability of the clusters, i.e. easy interpretation of the characteristics of the customers in the groups.
- The largest possible size of the smallest cluster, since economic rationality does not make it worthwhile to design a UI variant for the small number of customers.

A set of three metrics has been proposed to address these requirements, but it can be freely expanded in the future and adapted to meet other business needs:

- CSI is calculated according to the following formula:

$$CSI = \frac{CS_{max} - CS_{min}}{CS_{max}} \tag{13}$$

  where CS_{max} is the number of customers in the largest cluster, and CS_{min} is the number of customers in the smallest cluster.
- AAD is calculated according to the formula:

$$AAD = \frac{1}{k-1} \sum_{l=1}^{k-1} (A_{l+1} - A_l) \tag{14}$$

  where k is the number of outcome clusters, A_l is the average number of actions in the jth cluster, assuming that the clusters are ordered in descending order of the average number of actions in the online store.

- SSC is an absolute number representing the size (in terms of number of clients) of the smallest cluster.

These business-context metrics are designed to guide the selection of data clustering methods that are not only statistically sound but also practically relevant and economically viable for developing and maintaining distinct UI variants. For instance, a high SSC is desirable as it ensures that even the smallest user segment is large enough to justify a dedicated UI personalization effort, which is a direct business consideration.

The proposed list of metrics derived from the business context can be supplemented or modified as needed (e.g., the AAD metric based on the average number of customer actions can be replaced with metrics based on the average number of events, purchase values, returns, etc.).

The research conducted was divided into three parts:

1. Data clustering using GMM, GMM+, Fuzzy K-means, K-means, and BIRCH, for the number of clusters in the range from 4 to 8.
2. Application of the proposed adaptation of PROMETHEE II to select the optimal data clustering method, for the selected business context based on the described set of decision criteria.
3. Sensitivity analysis of the decisions resulting from the proposed approach according to the weights of the groups of decision factors.

The results are presented in the following subsections of the next chapter.

4. Results

4.1. Data clustering

Implementations of K-means, GMM, and BIRCH from the scikit-learn [87] library were used. Fuzzy K-means is implemented in the skfuzzy [88] library. Data manipulations during loading and preprocessing were performed using the NumPy and pandas libraries. The calculations were performed on a 3.2 GHz CPU with 32 active logical cores, as some algorithms use multithreaded computing technology. The algorithms had 128 GB of DDR5 RAM at their disposal. The algorithm implementations used the latest libraries within the Python language in version 3.11. Each algorithm was applied to the same data, with the parameter responsible for the number of clusters varying from 4 to 8. The comparison of the results is based on the final cluster assignment, but also includes various statistics collected during the computation and evaluation of the data clustering algorithms. In practice, metrics such as CPU and RAM usage need to be considered. Another important metric is time, especially when divided into training and evaluation parts since different algorithms handle it differently.

The dataset was filtered and aggregated based on user activity to create clusters. Each user was represented as a collection of activities condensed into a set of features. This representation converted the website activity logs into a feature space in which individual users correspond to unique data points. The dimensions of this space are derived from predetermined features that contain relevant information for future use, such as categories of frequently viewed products. This data representation shows multiple distributions, including numerical data that represent both near-continuous occurrences as well as quantities that occur infrequently. In addition, binary data has been added to uniquely encode the presence of certain features. Since common data clustering algorithms are sensitive to a scale of certain dimensions, the previous approach would overestimate some features. To unify the representation, additional summarization has been performed, i.e. extraction of the most common statistics (min, max, median, mean, standard deviation, length), as well as the Yeo-Johnson and quantile transforms. The final preprocessing is obtained after applying Principal Component Analysis (PCA) to remove strong correlations and reduce dimensionality. We selected principal components responsible for explaining more than 99% of the data variance.

Mini-Batch GMM significantly outperformed the standard GMM in terms of training speed and resource usage, demonstrating its suitability for large-scale or resource-constrained environments. The resource savings are especially noticeable at the level of RAM usage, as the use of mini-batches quadratically reduces memory requirements. However, while Mini-Batch GMM achieved comparable clustering results on the training set, its performance did not generalize well to the test set, as confirmed by all the metrics examined. This suggests that the model's estimates, although efficient, may suffer from reduced stability and accuracy due to the stochastic nature of mini-batch updates. In contrast, both the standard GMM and GMM+, despite higher computational cost, produced more robust and generalizable results and proved that, unlike Mini-Batch GMM, they are suitable for practical use in this case.

The application of the GMM+ algorithm in this study represents a groundbreaking technological advancement, offering a unique approach that has never been utilized before in similar applications. When compared to the traditional GMM method, GMM+ demonstrated equal or superior performance in most tests. Particularly in scenarios involving similar distributions, where the parameters converged to identical or nearly identical extremes, GMM+ proved to be faster, requiring fewer computations to reach convergence within the same number of steps. While some tests did show that GMM+ necessitated more steps than the default GMM, leading to longer computation times, a closer examination revealed a key distinction: the parameters derived from GMM+ were substantially different from those obtained using GMM. This divergence ultimately resulted in better data clustering, as measured by accepted criteria. This finding underscores the exceptional regularization capabilities of the GMM+ algorithm, especially when applied to real-world datasets.

GMM+ excels when the dataset contains redundancy, such as clusters of nearly identical points or smooth continuous data. Granulation reduces these without losing much information, enabling accurate yet faster learning. Because it approximates the full dataset more faithfully (by summarizing parts of it into granules), GMM+ achieves better test set performance than Mini-Batch GMM. GMM+ can approximate the behavior of full-batch GMM, preserving model quality, while avoiding the full cost. Due to these limitations, further research was limited to testing GMM and GMM+, dropping Mini-Batch GMM.

Training the BIRCH algorithm on the full dataset required more memory than was currently available, so it was done by sampling the original dataset. The memory consumption increases by leaps and bounds, but the results remain similar and change continuously with the size of the data used. For the remaining algorithms, the entire dataset was used for training.

Figs. 3 and 4 compare the clustering after projection on the first 4 PCA components (the diagonal of the graph indicates the density estimate of the data values projected along the given dimension). Fig. 4 shows that the individual clusters differ significantly more in size and shape (diagonal of the graph) than in the case of K-means. This representation is more accurate concerning the nature of the data, but results in significantly worse values for some data clustering metrics. K-means is suboptimal because it is limited to clusters with similar variance. GMM overcomes this disadvantage by estimating the variance for each cluster separately. Another implication of the graph is based on the distributions of each dimension: "pointed" distributions indicate large concentrations of similar observations around cluster centers. This can be exploited in GMM+ by combining similar observations into a single data granule, which speeds up the computations.

Figs. 5 and 6 show examples of clustering results and their format. It is worth noting that resource usage (Fig. 5) is not monotone with respect to the number of clusters since optimization problems formulated for data clustering algorithms can converge to different local extremes.

When analyzing the resulting indicator values, it should also be noted that some of the metrics may favor certain data clustering methods (for example, the CH index is based on a loss function used in K-means optimization). It is important to consider such relationships when selecting or weighting decision criteria.

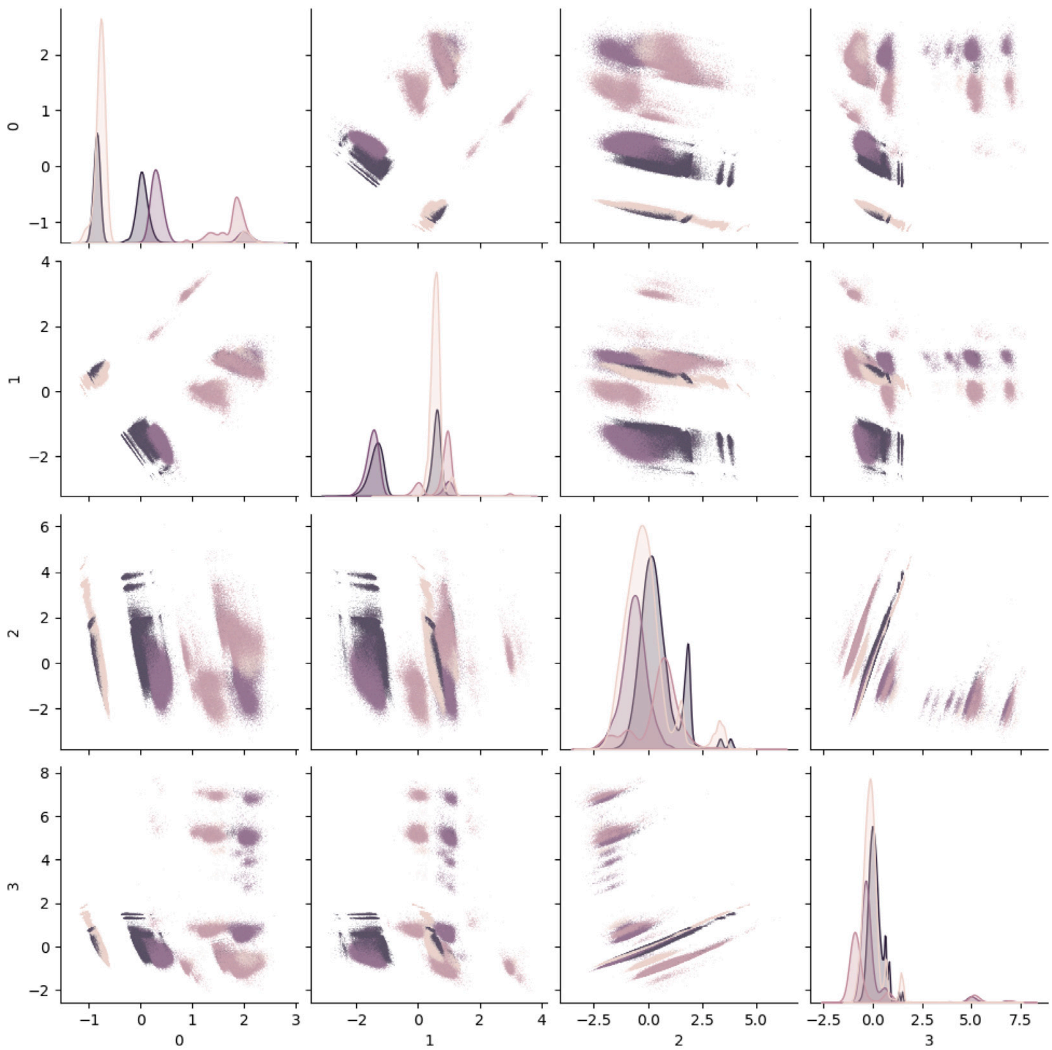

Fig. 3. Pairwise scatterplot of the first 4 PCA components of the dataset with labels produced by K-means for 4 different clusters.

4.2. Selection of the optimal data clustering method

Before starting the analysis, assumptions were made about the weights of the decision factor groups and the factors themselves. It was assumed that the weights would be equal so as not to favor any group or factor. However, the proposed approach does not preclude the use of different weights, in particular a weight of 0, which means excluding a group of factors or a single factor from the decision process. Evaluation of the impact of the change of weights was part of a sensitivity analysis of the ranking of data clustering methods described in the next subsection.

According to the proposed approach, in the first phase, the ranking of alternatives was calculated separately for each value of k. The basis for the calculation was a matrix of indicator values, an example of which (for $k = 6$) is shown in Table 2.

Analogous matrices were created for each value of k from 4 to 8.

The next step was to normalize the matrix of alternatives and decision criteria. The normalized decision matrix (Table 3) reveals that K-means frequently achieves optimal performance (normalized value of 1) on individual context-free metrics. This is an expected outcome given that K-means' objective function directly optimizes characteristics like the intra-cluster sum of squares, which correlate with metrics such as a lower Davies–Bouldin Index or a higher Silhouette Score under certain data distributions. However, its potential weaker performance on specific business-context or resource-usage metrics may highlight the necessity of a multi-criteria approach, as singular metric optimality does not guarantee overall suitability.

In multi-criteria analysis, however, this is not a determinant of the final ranking of alternatives. To calculate it, it is necessary to compare all pairwise, for each decision factor separately.

The next step is to compute a matrix containing the values of the preference function for each value of the parameter k and for each decision criterion. An example of such a matrix for $k = 6$ and the decision criterion *Silhouette Score* is shown in Table 4. In addition, equal weighting of groups of decision factors and equal weighting of criteria within a group were considered in the calculations.

The results of all decision factors are aggregated by calculating the positive (ϕ^+) and negative (ϕ^-) outranking flows. This operation is performed independently for each tested k value, and its example result (for $k = 6$) is shown in Table 5.

The last computational element in the first part of the proposed adaptation of the PROMETHEE II method is the calculation of the final outranking flows (independently for each k value), which is the basis for determining the ranking of the alternatives. The calculated values for each k analyzed are presented in Table 6. For $k = 6$, GMM+ ($\phi = 0.202$) emerged as the top-ranked algorithm, marginally outperforming GMM ($\phi = 0.184$) and K-means ($\phi = 0.195$). A closer examination of its performance profile across criteria groups indicates that GMM+'s advantage likely stems from a more balanced performance, particularly excelling in business-context metrics due to its regularization effect,

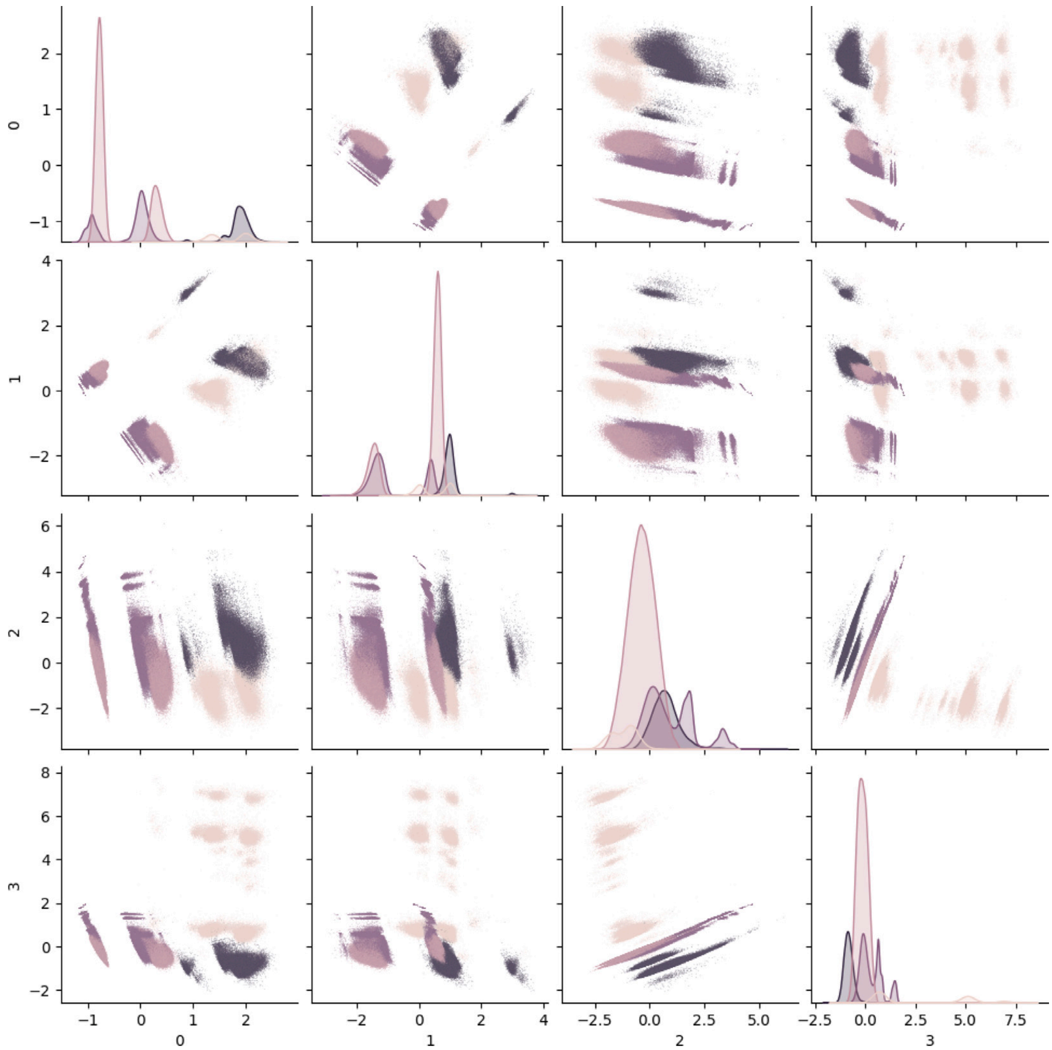

Fig. 4. Pairwise scatterplot of the first 4 PCA components of the dataset with labels produced by GMM+ for 4 different clusters.

	n	train_time	train_cpu	train_mem	train_mem_max
0	4	10.345467	0.632558	66466764	290138182
1	5	11.847418	0.634359	70277268	315100416
2	6	11.472513	0.616842	71964243	331782419
3	7	29.626010	0.655140	71321798	337539523
4	8	13.966047	0.595801	72244300	352160204

Fig. 5. The summary of time and resources used during the training phase of the GMM+ algorithm for different numbers of clusters (parameter n).

leading to more uniform cluster sizes (SSC) or better activity diversity (AAD), while maintaining competitive scores in context-free quality. This suggests that its granule-based approximation may offer benefits in handling the heterogeneity of real-world e-commerce data, leading to a more practically useful segmentation according to our comprehensive criteria set.

Conversely, Fuzzy K-means consistently ranked lower (e.g., ϕ = -0.375 for k = 6), potentially because the inherent fuzziness did not align well with the defined evaluation criteria or data structure and the higher computational demands (CT), which outweighed its scores in other areas.

The alternatives are ranked according to the computed ϕ value, in descending order. The highest value for each k is highlighted in bold in the table, as it indicates the best option for a given number of resulting clusters. A complete summary of the rankings for each data clustering algorithm is shown in Fig. 7.

Analyzing the results of this part of the proposed approach, it can be seen that the K-means algorithm could be considered the best option

	0	1	2	3	4
silh	0.119736	0.100149	0.093643	0.100829	0.090585
ent	0.777268	0.656807	0.649166	0.638929	0.768725
ch	44300.043043	40510.903191	44108.211349	46443.940158	47606.817351
db	2.917541	2.235446	2.084676	1.78964	2.095044
dunn	0.019366	0.06695	0.079705	0.077123	0.021704
clusters	[47079, 483320, 176343, 111410]	[35796, 15055, 119429, 120826, 527046]	[23038, 26954, 110322, 120918, 512602, 24318]	[23364, 21166, 499191, 99815, 129734, 24314, 2...	[25035, 33652, 236223, 86054, 24312, 11950, 31...
time	0.512609	0.689954	0.968645	1.062735	1.279971
cpu	0.8478	0.786527	0.707474	0.744568	0.730411
mem	6553760	6555936	6558040	6560096	6562700
mem_max	320793448	327340648	333888144	340435192	346983188

Fig. 6. The summary of data clustering metrics (Silhouette Coefficient, scaled entropy, Caliński–Harabasz index, Davies–Bouldin index, Dunn index), sizes of different clusters, and resources used during the evaluation phase of the GMM+ algorithm.

Table 2
Matrix of indicator values for $k = 6$.

Algorithm	Context-free			Resource usage			Business context		
	SS	DBI	DI	AC	MM	CT	SSC	CSI	AAD
GMM	1.094	2.084	0.071	0.775	3.338	0.884	0.028	0.955	0.091
GMM+	1.094	2.085	0.080	0.707	3.338	0.969	0.028	0.955	0.091
K-means	1.112	1.834	0.112	0.947	1.080	0.075	0.025	0.950	0.021
Fuzzy K-means	0.991	4.514	0.002	0.051	3.600	0.488	0.004	0.992	0.028
BIRCH	1.095	2.209	0.019	0.697	10.795	3.773	0.028	0.950	0.002
Min	0.991	1.834	0.002	0.051	1.080	0.075	0.004	0.950	0.002
Max	1.112	4.514	0.112	0.947	10.795	3.773	0.028	0.992	0.091

Table 3
Normalized decision matrix for $k = 6$.

Algorithm	Context-free			Resource usage			Business context		
	SS	DBI	DI	AC	MM	CT	SSC	CSI	AAD
GMM	0.851	0.907	0.625	0.192	0.767	0.781	0.988	0.877	0.998
GMM+	0.846	0.907	0.708	0.267	0.767	0.758	0.988	0.878	1.000
K-means	1.000	1.000	1.000	0.000	1.000	1.000	0.860	0.995	0.210
Fuzzy K-means	0.000	0.000	0.000	1.000	0.741	0.888	0.000	0.000	0.295
BIRCH	0.854	0.860	0.154	0.279	0.000	0.000	1.000	1.000	0.000

Table 4
Matrix of weighted preference function values for $k = 6$ and Silhouette Score.

	GMM	GMM+	K-means	Fuzzy K-means	BIRCH
GMM		**0.001**	0.000	**0.095**	0.000
GMM+	0.000		0.000	**0.094**	0.000
K-means	**0.017**	**0.017**		**0.111**	**0.016**
Fuzzy K-means	0.000	0.000	0.000		0.000
BIRCH	0.000	**0.001**	0.000	**0.095**	

Table 5
The positive (ϕ^+) and negative (ϕ^-) outranking flows for $k = 6$.

	GMM	GMM+	K-means	Fuzzy K-means	BIRCH	ϕ^+
GMM		0.003	0.123	0.553	0.341	0.255
GMM+	0.018		0.132	0.562	0.347	0.265
K-means	0.132	0.126		0.581	0.371	0.302
Fuzzy K-means	0.102	0.096	0.120		0.294	0.153
BIRCH	0.025	0.017	0.047	0.430		0.130
ϕ^-	0.069	0.060	0.106	0.531	0.338	

three times (for $k = 4, 5, 8$) and the GMM+ algorithm twice (for $k = 6, 7$).

In this part of the study, the combinations of the data clustering algorithm and the number of resulting clusters that turned out to have the best rankings for specific k values appear as alternatives in the decision matrix. In this case, this means that three alternatives applying the K-means algorithm and two alternatives using the GMM+ algorithm will be compared. The best options from the first part of the survey are compared, and the results do not include rejected options that ranked lower in previous rankings.

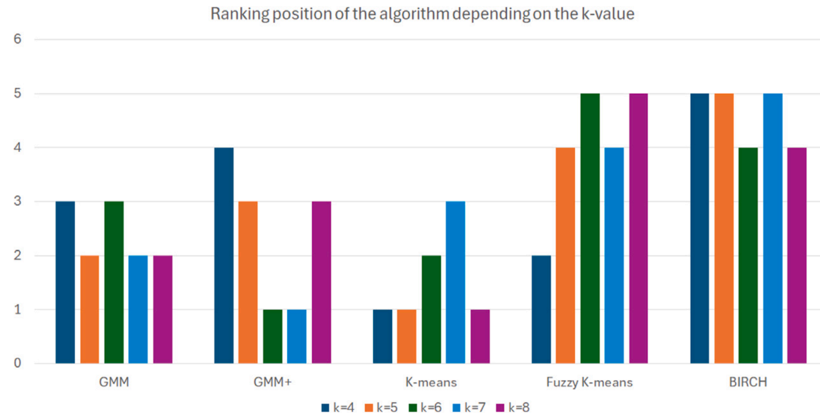

Fig. 7. Ranking position of the algorithm depending on the k value.

Table 6
The final outranking flows (ϕ values).

	k = 4	k = 5	k = 6	k = 7	k = 8
GMM	−0.074	0.046	0.184	0.221	0.073
GMM+	−0.096	−0.014	**0.202**	**0.224**	−0.047
K-means	**0.224**	**0.300**	0.195	0.070	**0.397**
Fuzzy K-means	0.053	−0.149	−0.375	−0.211	−0.291
BIRCH	−0.108	−0.182	−0.207	−0.304	−0.132

Table 7
Ranking of the best alternatives.

Rank	Alternative	ϕ value
1	K-means (k = 4)	0.252
2	K-means (k = 5)	0.094
3	GMM+ (k = 6)	−0.106
4	K-means (k = 8)	−0.108
5	GMM+ (k = 7)	−0.133

The results of the calculations (Table 7) indicate that, given the initial assumptions made (in particular, regarding the equal weights of the groups of decision factors and the criteria themselves), the optimal method for clustering behavioral data for serving dedicated UI variants in e-commerce would be the K-means algorithm and $k = 4$. The final “best of the best” ranking identifies K-means with $k = 4$ as the optimal choice ($\phi = 0.252$) under our initial assumption of equal criteria group weights. This outcome reflects K-means’ ability, at this lower k value, to achieve a strong balance across the diverse criteria. Specifically, for $k = 4$, it likely provided a good compromise between forming distinct, meaningful clusters (reflected in context-free and business-context metrics like SSC) without incurring excessive computational cost or overly fragmenting the user base, which can happen at higher k values. The clear advantage over the next ranked option (K-means $k = 5$, $\phi = 0.094$) suggests robustness in this choice for the given conditions.

The advantage of the best alternative over the next ranked options is clear, which lets to believe that the result is reliable. An additional verification of such a recommendation is a sensitivity analysis, which is described in the next subsection.

It should be noted that the alternative indicated by the proposed extension of the PROMETHEE II method coincides with the expert decision made for the study of the effectiveness of dedicated UI options [2, 37].

4.3. Sensitivity analysis

The purpose of sensitivity analysis is to examine the impact of assumptions made on the results obtained. This phase of the research was divided into two parts:

- Analysis of the effect of the weights on the recommendation of the best alternative for a fixed k value.
- Analysis of the effect of the weights of different combinations of decision factor groups on the recommendation of the “best of the best” option.

In the first case, the focus was on calculations for $k = 6$. The original assumption of equality of weights of all groups of decision factors was replaced by an assumption according to which performance factors were excluded (their weight was set to 0), and the proportion between context-free and business context factor groups could vary from 0 to 1 (with the rule that the sum of weights is equal to 1).

A graphical representation of such an analysis is shown in Fig. 8.

It can be seen that the methods of the GMM family are clearly superior when the business context criteria are given a high weight. As the proportion of context-free metrics increases, their advantage decreases. From a context-free weight of 0.6, the K-means algorithm would be considered the best option. It reveals a critical sensitivity for $k = 6$: GMM and GMM+ are superior when business-context criteria (e.g., cluster balance, activity diversity, minimum size) are prioritized. This superiority is likely due to their probabilistic nature and, for GMM+, its regularization, allowing for more flexible cluster shapes and sizes that better align with these business-driven objectives. Conversely, when context-free metrics (like Silhouette Score or DBI, at which K-means often excels due to its objective function) are weighted more heavily, K-means becomes the preferred method. This crossover analytically demonstrates that the “optimal” algorithm is highly dependent on strategic business priorities.

It is also worth noting that the GMM+ algorithm produces better results than GMM regardless of the weights of the decision factor groups, which shows that this variant is an improvement over the basic solution.

In the second case, the focus was on selecting the best alternative. This time, the sensitivity analysis was performed in three versions:

- The impact of the relationship between context-free factors and criteria derived from business requirements, while ignoring the set of performance factors (Fig. 9).
- The impact of the relationship between context-free factors and resource-usage criteria, while ignoring the set of business-context factors (Fig. 10).
- The impact of the relationship between resource-usage criteria and business-context factors, while ignoring context-free indicators (Fig. 11).

To summarize this part of the sensitivity analysis, it is worth noting that regardless of the combination of the weights of the groups of decision factors, the best alternative always remained the solution using the K-means algorithm, with only the k value changing. This indicates

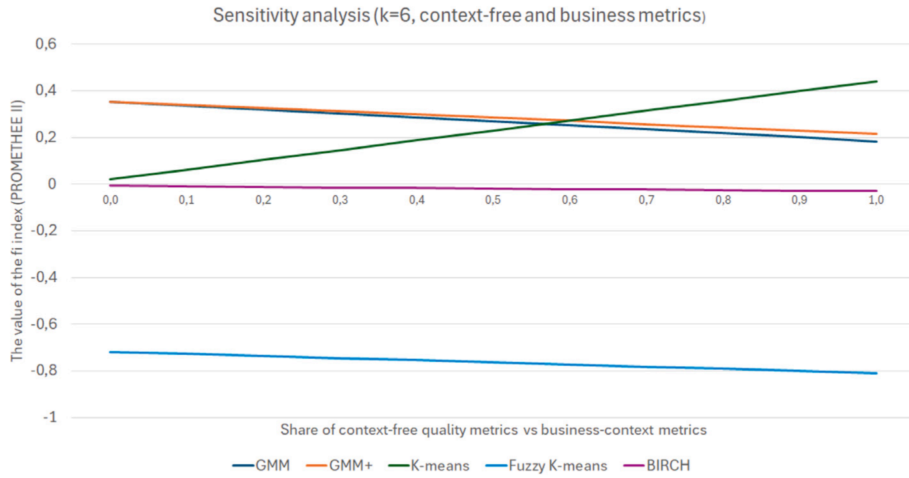

Fig. 8. Sensitivity analysis ($k = 6$, context-free and business metrics).

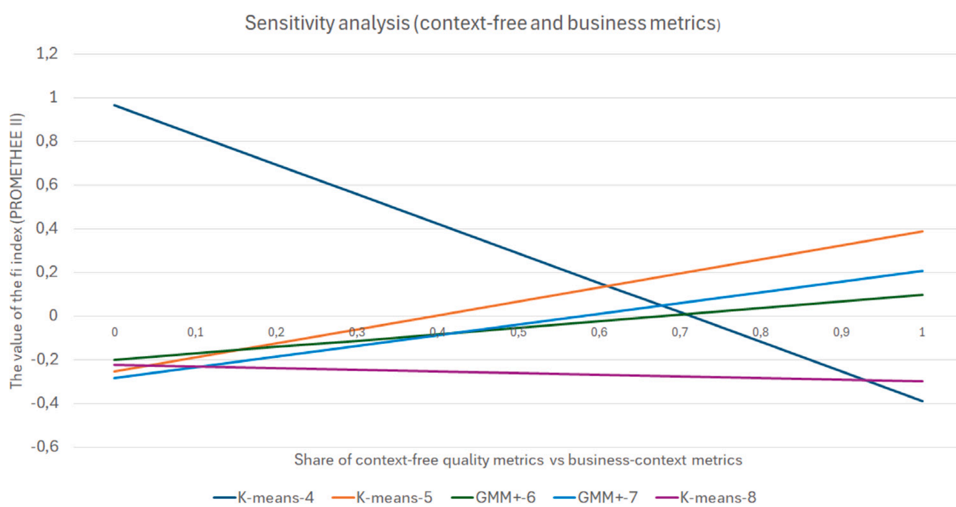

Fig. 9. Sensitivity analysis (context-free and business metrics).

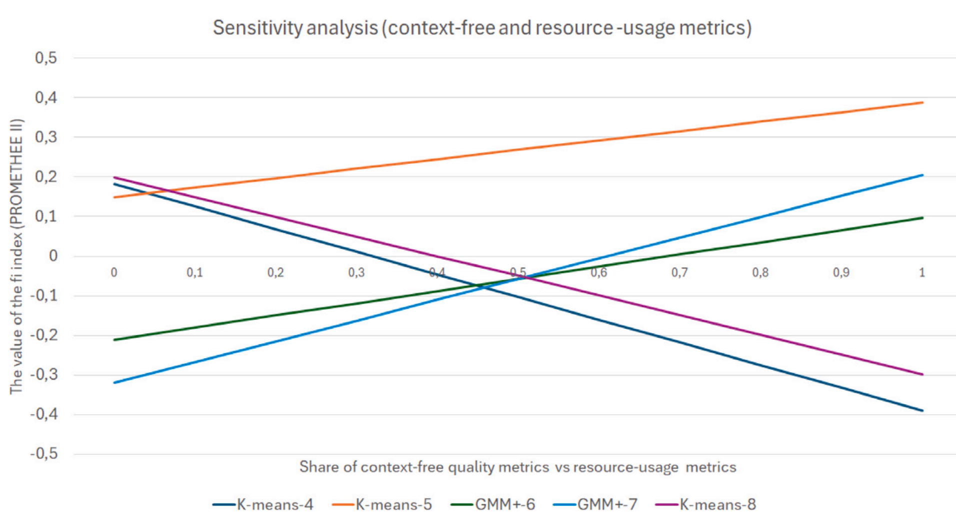

Fig. 10. Sensitivity analysis (context-free and resource-usage metrics).

that in the business context described above, K-means can be treated as the primary way of clustering customers when economic or technical considerations do not allow the development of a solution that implements multiple data clustering algorithms. The sensitivity analyses for the ‘best of the best’ selection demonstrate remarkable robustness in the preference for K-means-based solutions. For instance, Fig. 9 shows that even as the balance shifts between context-free and business-context criteria (ignoring resource usage), K-means (primarily $k = 4$

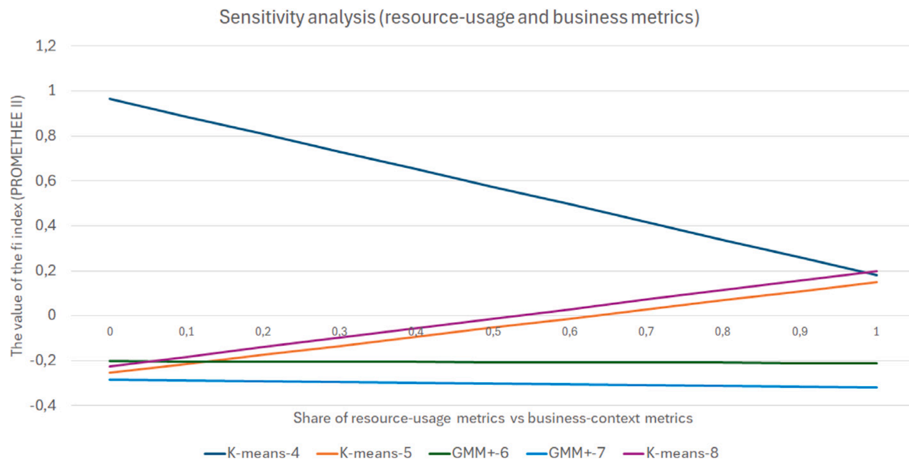

Fig. 11. Sensitivity analysis (resource-usage and business metrics).

or $k = 5$) consistently outperforms GMM+ options. This suggests that K-means's overall profile across these two crucial dimensions offers a more consistently superior trade-off for this dataset than GMM+, whose advantages might be more pronounced under specific conditions. The shift in optimal k for K-means (e.g., from $k = 4$ to $k = 5$ or $k = 8$ under different scenarios) indicates that while the algorithm choice is stable, the ideal segmentation granularity is indeed sensitive to which criteria group (e.g., resource usage vs. business context in Fig. 11) is de-emphasized.

5. Discussion

5.1. Summary of the conducted experimental studies

The analysis performed allows us to answer the research questions posed. The main observation, related to question **RQ.1**, significantly aided by our novel two-step adaptation of PROMETHEE II, which allowed for comparing 'best of the best' options across different k values, confirms that the ranking of clustering options depends on the predefined number of resulting clusters (k). Moreover, only for one combination of decision factors (equal share of all groups of metrics and measures included in them) was the ranking the same for two values of k ($k = 4$ and $k = 5$). For all other combinations, the rankings were different. Given the business context of the study, this fact should be taken into account when deciding to design a certain number of UI variants, as it may result in the need to change the way customers are grouped. However, if there is uncertainty about how many UI variants to design, one can use the second part of the proposed adaptation of the PROMETHEE II method and perform a ranking to find the "best of the best" alternative. This is particularly important when the differences in the rankings concern the first positions, which can be taken as a basic recommendation for the choice of the clustering algorithm (as was the case in the presented study). In such a situation, the choice of a clustering method and its initialization parameters can be difficult, and the proposed adaptation of PROMETHEE II makes it easier.

The analysis of the recommendations obtained after applying the MCA method allows us to conclude that they are generally in line with the expectations based on the specificities of the clustering algorithms analyzed and the expert knowledge, which corresponds to **RQ.2**. In addition, they confirm the validity of the assumptions made for the implementation of the research related to the business effectiveness of the mechanism of multivariate UI in e-commerce [2] because the applied MCDM recommended those methods that were considered best based solely on expert knowledge. At the same time, it is worth noting that the described research also confirmed doubts about which method (K-means or GMM+) would be better. The results of the formal multi-criteria analysis also gave different recommendations depending on the value of the k parameter.

Analysis of the results obtained provides a basis for answering research question **RQ.3**. The insights gained regarding this aspect, particularly the impact of business-context criteria, underscore the value of our original framework that explicitly incorporates such tailored metrics, moving beyond purely technical evaluations. First, it turns out that the alternative designated as "best of the best" may not be so if assumptions about the weights of groups of decision factors change. This means that depending on the set of criteria and their weights, different data clustering method recommendations can be obtained. This awareness should also accompany the selection of a data clustering method for multivariant UI in e-commerce, as it involves the possibility of giving higher priority to more important decision criteria (e.g., related to the fit with the business context). From another perspective, performance issues can be highly emphasized when constraints are related to the performance of available servers or the size of the dataset. Our sensitivity analysis provides a systematic way to explore these contingencies, offering an advancement over studies that present rankings based on single, fixed weighting schemes. This highlights how our framework allows for a more nuanced understanding than what is typically derived from existing literature, enabling decision makers to align method selection more closely with specific business goals.

It is worth noting that only a predefined range of values of the k parameter has been analyzed. This decision was based on the limitations of the resulting context and assumptions about serving dedicated UI variants. If customers were divided into too many clusters, the number of users in some of them might be too small to make it economically viable to design a dedicated layout variant for them. The set maximum number of clusters is a limitation of the research conducted, but the use of MCDM provides an opportunity to address the problem of the minimum number of users in clusters in a different way. Since the proposed decision factors include those related to the number of customer segments, they can be given a higher weight to emphasize the importance of such criteria. This type of analysis, including extended studies of the sensitivity of recommendations to the weights assigned to measures of clustering quality, will be the subject of future research.

5.2. Practical applications

The proposed approach, which integrates multi-factor evaluation of clustering quality, can be applied in various business contexts. It allows for a comprehensive consideration of different characteristics of clustering algorithms, which is often not the case in practical applications. Typically, decisions are made on the basis of single factors,

Table 8
Recommendations for single groups of decision making factors.

k/group	Context-free	Resource-usage	Business-context
4	BIRCH	Fuzzy K-means	K-means
5	K-means	Fuzzy K-means	K-means
6	K-means	Fuzzy K-means	GMM+
7	GMM+	Fuzzy K-means	GMM
8	K-means	Fuzzy K-means	K-means

which makes them not necessarily optimal. Evidence for this claim was provided by a sensitivity analysis of the recommendations to the combination of indicator groups and to the clustering quality measures themselves.

In the analyzed business context, limiting the decision process to a single metric (or even a single group of metrics) could significantly affect the recommendation for selecting the optimal clustering method. Recommendations resulting from considering only one set of metrics are shown in Table 8. These results show that different decisions could be expected depending on the choice of metrics group. Particularly noteworthy is the fact that if only resource usage is considered, Fuzzy K-means would be identified as the optimal algorithm, while the presented multi-criteria analysis would not recommend this approach.

The advantages of using multivariate analysis mean that the approach presented can be widely applied both in e-commerce and beyond. In the first case, it can be used in any application that requires customer segmentation or profiling for targeted communications, personalized offers, or pricing. In each of these areas, the right choice of grouping method is critical to business efficiency and determines success or failure. Therefore, the limitations of decisions based on single factors and the risks associated with a suboptimal decision should drive decision makers to use MCDM. The proposed adaptation of PROMETHEE II further addresses the specificities of the decision process where the number of clusters is not precisely defined. By developing the basic approach, decision makers not only receive a recommendation for a clustering algorithm but can also compare the best options for each value of the k parameter.

The relevance of the presented results is not limited to e-commerce. The presented concept can be used in any business context where different aspects of clustering quality need to be considered in the decision making process. The only limitation is the set of decision factors. While the groups of metrics defined as “context-free” and “resource-usage” are universal and can be used regardless of the application area, the definition of business context-related metrics must be tailored accordingly. Although this is an additional workload, it can yield tangible benefits by providing a comprehensive view of the strengths and weaknesses of different clustering algorithms.

6. Conclusion

This paper presents a significant advancement in the methodology for selecting data clustering algorithms for e-commerce personalization by introducing a novel adaptation of the PROMETHEE II method and a comprehensive, business-aware evaluation framework. By integrating both technical and business criteria, it provides a comprehensive assessment that can be successfully applied in practice. The results described provide insight into the usefulness of different data clustering algorithms for improving user experience through tailored interactions and can serve as a valuable resource for e-commerce practitioners and researchers, helping them choose the most effective grouping methods to achieve excellent personalization and ultimately higher customer satisfaction and better retention. One of the main original contributions of this study is the innovative application of the PROMETHEE II method, which offers a structured and robust solution for selecting optimal clustering strategies (both algorithms and the number of clusters) in scenarios where business requirements for segmentation granularity are not fixed.

Some notable findings that uniquely contribute to the field and address limitations of previous studies include:

- Businesses should consider multiple data clustering methods and, using various decision criteria derived from the characteristics of data clustering algorithms, select the one that best fits their specific data and personalization goals.
- Implementing a multi-criteria evaluation framework that includes both technical and business characteristics can maximize the effectiveness of customer segmentation efforts and ensure that grouping results are not only formally valid, but also practically relevant to increasing business value.
- Since the dynamic nature of e-commerce requires constant monitoring and improvement, it is critical to regularly review and update the segmentation methodology to reflect changing customer behaviors and preferences.
- An interesting alternative to classic clustering techniques is the GMM+ approach, which has twice been among the best options obtained in the decision process.

However, the limitations observed during the study cannot be overlooked. Undoubtedly, raw data plays a key role in any kind of data clustering. The paper presents results based on the customer behavior of an online store. While the proposed adaptation of the PROMETHEE II method is universal and can be applied regardless of the user data, to draw general conclusions about specific data clustering algorithms, it would be necessary to repeat the study on customers of other e-shops. It is also worth mentioning that the analysis covered only five data clustering algorithms, although more methods could be used. Therefore, further studies should be considered to verify other techniques, their modifications, and extensions. The described example of GMM and GMM+ shows that the proposed approach can also be used to comprehensively evaluate improvements to existing algorithms, providing a valuable tool for validating new concepts and ideas. The importance of the choice of feature weights and decision criteria should also not be overlooked. The sensitivity analysis described highlighted this issue, and the question of appropriate prioritization of decision criteria should be studied in depth. Finally, it is important to keep in mind that the paper uses only one method of multi-criteria decision making, while there are many more approaches. It cannot be excluded that the rankings obtained after applying another method (e.g. ELECTRE) would lead to different recommendations. This is also an interesting direction for further research. Moreover, the range of k = 4-8 was chosen based on practical considerations for e-commerce UI variant design. While our framework could be applied to a wider range, exploring very high values of k would require careful consideration of diminishing returns in segment distinctiveness and the economic feasibility of managing numerous UI variants. Our business-context metrics, particularly SSC, already aim to penalize overly granular segmentations. Future research could investigate the sensitivity of business metrics to a broader range of k values.

In conclusion, the multi-criteria evaluation of data clustering methods presented in this article highlights the key role of customer segmentation in the development of effective personalized e-commerce user interfaces. By balancing technical performance with business relevance, the proposed comprehensive evaluation framework provides valuable guidance for algorithm selection and customer experience improvement. Ultimately, the insights gained from this study can help e-commerce platforms achieve higher customer satisfaction, greater engagement, and long-term business success. The development of a comprehensive multi-criteria evaluation framework, uniquely integrating technical, resource, and specifically designed e-commerce business context metrics, represents a significant step towards more practically relevant algorithm selection in e-commerce applications.

CRediT authorship contribution statement

Elżbieta Pawełek-Lubera: Project administration. **Mateusz Przyborowski:** Writing – original draft, Software, Data curation. **Dominik Ślęzak:** Writing – review & editing, Validation. **Adam Wasilewski:** Methodology, Formal analysis, Conceptualization, Writing – original draft.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Elzbieta Pawelek-Lubera reports financial support was provided by National Centre for Research and Development. Other authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The behavioral data was collected using the AIM^2 platform developed by Fast White Cat S.A., Poland under the project "Self-adaptation of the online store interface for the customer requirements and behaviour" co-funded by the National Centre for Research and Development, Poland under the Sub-Action 1.1.1 of the Operational Programme Intelligent Development 2014–2020.

Data availability

Data will be made available on request.

References

[1] A. Hiziroglu, Soft computing applications in customer segmentation: State-of-art review and critique, Expert Syst. Appl. 40 (16) (2013) 6491–6507, http://dx.doi.org/10.1016/j.eswa.2013.05.052.
[2] A. Wasilewski, M. Przyborowski, Clustering methods for adaptive e-commerce user interfaces, in: Rough Sets: International Joint Conference, IJCRS 2023, Krakow, Poland, October 5–8, 2023, Proceedings, Springer-Verlag, Berlin, Heidelberg, 2024, pp. 511–525, http://dx.doi.org/10.1007/978-3-031-50959-9_35.
[3] M. Przyborowski, D. Slezak, Approximation of the expectation-maximization algorithm for Gaussian mixture models on big data, in: 2022 IEEE International Conference on Big Data, Big Data, 2022, pp. 6256–6260, http://dx.doi.org/10.1109/BigData55660.2022.10020450.
[4] V. Salonen, H. Karjaluoto, Web personalization: The state of the art and future avenues for research and practice, Telemat. Inform. 33 (4) (2016) 1088–1104, http://dx.doi.org/10.1016/j.tele.2016.03.004.
[5] M.F. Dzulfikar, B. Purwandari, D.I. Sensuse, J.S. Lusa, I. Solichah, P. Prima, I. Wilarso, Personalization features on business-to-consumer e-commerce: Review and future directions, in: 2018 4th International Conference on Information Management, ICIM, 2018, pp. 220–224, http://dx.doi.org/10.1109/INFOMAN.2018.8392839.
[6] J. Guo, W. Zhang, T. Xia, Impact of shopping website design on customer satisfaction and loyalty: The mediating role of usability and the moderating role of trust, Sustainability 15 (8) (2023) http://dx.doi.org/10.3390/su15086347.
[7] O. Saoula, A. Shamim, N. Mohd Suki, M. Javed, M. Abid, A. Patwary, A. Abbasi, Building e-trust and e-retention in online shopping: the role of website design, reliability and perceived ease of use, Span. J. Mark. - ESIC 27 (2023) http://dx.doi.org/10.1108/SJME-07-2022-0159.
[8] G. Hassna, K. Rouibah, P. Benjamin Lowry, J. Paliszkiewicz, M. Mądra-Sawicka, The roles of user interface design and uncertainty avoidance in B2C ecommerce success: Using evidence from three national cultures, Electron. Commer. Res. Appl. 61 (2023) 101297, http://dx.doi.org/10.1016/j.elerap.2023.101297.
[9] D. Desai, An Empirical Study of Website Personalization Effect on Users Intention to Revisit E-Commerce Website Through Cognitive and Hedonic Experience: Proceedings of ICDMAI 2018, Volume 2, 2019, pp. 3–19, http://dx.doi.org/10.1007/978-981-13-1274-8_1.
[10] A. Mantha, A. Sundaresan, S. Kedia, Y. Arora, S. Gupta, G. Wang, P. Kanumala, S. Guo, K. Achan, A real-time whole page personalization framework for E-commerce, in: 2020 IEEE International Conference on Big Data, Big Data, 2020, pp. 4646–4650, http://dx.doi.org/10.1109/BigData50022.2020.9378254.
[11] S. Roychowdhury, E. Alareqi, W. Li, OPAM: Online purchasing-behavior analysis using machine learning, in: 2021 International Joint Conference on Neural Networks, IJCNN, 2021, pp. 1–8, http://dx.doi.org/10.1109/IJCNN52387.2021.9533658.
[12] H. Hallikainen, M. Luongo, A. Dhir, T. Laukkanen, Consequences of personalized product recommendations and price promotions in online grocery shopping, J. Retail. Consum. Serv. 69 (2022) 103088, http://dx.doi.org/10.1016/j.jretconser.2022.103088.
[13] A. Khamaj, A.M. Ali, Adapting user experience with reinforcement learning: Personalizing interfaces based on user behavior analysis in real-time, Alex. Eng. J. 95 (2024) 164–173, http://dx.doi.org/10.1016/j.aej.2024.03.045.
[14] L. Theodorakopoulos, A. Theodoropoulou, Leveraging big data analytics for understanding consumer behavior in digital marketing: A systematic review, Hum. Behav. Emerg. Technol. 2024 (2024) http://dx.doi.org/10.1155/2024/3641502.
[15] M.H. Gouthier, C. Nennstiel, N. Kern, L. Wendel, The more the better? Data disclosure between the conflicting priorities of privacy concerns, information sensitivity and personalization in e-commerce, J. Bus. Res. 148 (2022) 174–189, http://dx.doi.org/10.1016/j.jbusres.2022.04.034.
[16] A. Alkis, T. Kose, Privacy concerns in consumer E-commerce activities and response to social media advertising: Empirical evidence from Europe, Comput. Hum. Behav. 137 (2022) 107412, http://dx.doi.org/10.1016/j.chb.2022.107412.
[17] H. Cai, A. Mardani, Research on the impact of consumer privacy and intelligent personalization technology on purchase resistance, J. Bus. Res. 161 (2023) 113811, http://dx.doi.org/10.1016/j.jbusres.2023.113811.
[18] A. Petrosyan, Global Number of Internet Users 2005–2023, Internet Report, Statista, 2024, https://www.statista.com/statistics/273018/number-of-internet-users-worldwide/. (Accessed 28 May 2024).
[19] B. Cooil, L. Aksoy, T.L. Keiningham, Approaches to customer segmentation, J. Relatsh. Mark. 6 (3–4) (2008) 9–39.
[20] G. Adomavicius, A. Tuzhilin, Expert-driven validation of rule-based user models in personalization applications, Data Min. Knowl. Discov. 5 (2001) 33–58.
[21] R.-S. Wu, P.-H. Chou, Customer segmentation of multiple category data in e-commerce using a soft-clustering approach, Electron. Commer. Res. Appl. 10 (3) (2011) 331–341.
[22] M. Gomes, T. Meisen, A review on customer segmentation methods for personalized customer targeting in e-commerce use cases, Inf. Syst. E- Bus. Manag. 21 (2023) 1–44, http://dx.doi.org/10.1007/s10257-023-00640-4.
[23] M. Kaptein, P. Parvinen, Advancing e-commerce personalization: Process framework and case study, Int. J. Electron. Commer. 19 (3) (2015) 7–33.
[24] T. Kansal, S. Bahuguna, V. Singh, T. Choudhury, Customer segmentation using K-means clustering, in: 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems, CTEMS, IEEE, 2018, pp. 135–139.
[25] Q. Su, L. Chen, A method for discovering clusters of e-commerce interest patterns using click-stream data, Electron. Commer. Res. Appl. 14 (1) (2015) 1–13.
[26] B. Lorbeer, A. Kosareva, B. Deva, D. Softić, P. Ruppel, A. Küpper, Variations on the clustering algorithm BIRCH, Big Data Res. 11 (2018) 44–53, http://dx.doi.org/10.1016/j.bdr.2017.09.002, Selected papers from the 2nd INNS Conference on Big Data: Big Data & Neural Networks.
[27] A. YYYSalamzadeh, P. Ebrahimi, M. Soleimani, M. Fekete-Farkas, Grocery apps and consumer purchase behavior: application of Gaussian mixture model and multi-layer perceptron algorithm, J. Risk Financ. Manag. 15 (10) (2022) 424.
[28] X. Zhang, X. Shen, T. Ouyang, Extension of DBSCAN in online clustering: An approach based on three-layer granular models, Appl. Sci. 12 (19) (2022) http://dx.doi.org/10.3390/app12199402.
[29] H. Singh, P. Kaur, An effective clustering-based web page recommendation framework for E-commerce websites, SN Comput. Sci. 2 (2021) http://dx.doi.org/10.1007/s42979-021-00736-z.
[30] J. Ke, Y. Wang, M. Fan, X. Chen, W. Zhang, J. Gou, Discovering e-commerce user groups from online comments: An emotional correlation analysis-based clustering method, Comput. Electr. Eng. 113 (2024) 109035, http://dx.doi.org/10.1016/j.compeleceng.2023.109035.
[31] R. Kuo, M.N. Alfareza, T.P.Q. Nguyen, Genetic based density peak possibilistic fuzzy c-means algorithms to cluster analysis- a case study on customer segmentation, Eng. Sci. Technol. Int. J. 47 (2023) 101525, http://dx.doi.org/10.1016/j.jestch.2023.101525.
[32] K. Zheng, X. Huo, S. Jasimuddin, J.Z. Zhang, O. Battaïa, Logistics distribution optimization: Fuzzy clustering analysis of e-commerce customers' demands, Comput. Ind. 151 (2023) 103960, http://dx.doi.org/10.1016/j.compind.2023.103960.
[33] K. Wang, T. Zhang, T. Xue, Y. Lu, S.-G. Na, E-commerce personalized recommendation analysis by deeply-learned clustering, J. Vis. Commun. Image Represent. 71 (2020) 102735, http://dx.doi.org/10.1016/j.jvcir.2019.102735.
[34] H. Khatter, S. Arif, U. Singh, S. Mathur, S. Jain, Product recommendation system for E-commerce using collaborative filtering and textual clustering, in: 2021 Third International Conference on Inventive Research in Computing Applications, ICIRCA, IEEE, 2021, pp. 612–618.
[35] Y. Gulzar, A.A. Alwan, R.M. Abdullah, A.Z. Abualkishik, M. Oumrani, OCA: Ordered clustering-based algorithm for E-commerce recommendation system, Sustainability 15 (4) (2023) http://dx.doi.org/10.3390/su15042947.
[36] D. Prajapati, A.R. Harish, Y. Daultani, H. Singh, S. Pratap, A clustering based routing heuristic for last-mile logistics in fresh food E-commerce, Glob. Bus. Rev. 24 (1) (2023) 7–20.
[37] A. Wasilewski, E. Pawełek-Lubera, Practical assessment of the e-commerce multivariant user interface, Inf. Softw. Technol. 181 (2025) 107684, http://dx.doi.org/10.1016/j.infsof.2025.107684.
[38] G. Mclachlan, K. Basford, Mixture Models: Inference and Applications to Clustering, Vol. 38, 1988, http://dx.doi.org/10.2307/2348072.
[39] C. Biernacki, G. Celeux, G. Govaert, Assessing a mixture model for clustering with the integrated completed likelihood, IEEE Trans. Pattern Anal. Mach. Intell. 22 (7) (2000) 719–725, http://dx.doi.org/10.1109/34.865189.

[40] J.-P. Baudry, A.E. Raftery, G. Celeux, K. Lo, R. Gottardo, Combining mixture components for clustering, J. Comput. Graph. Statist. 19 (2) (2010) 332–353.
[41] C. Maugis, G. Celeux, M.-L. Martin-Magniette, Variable selection for clustering with Gaussian mixture models, Biometrics 65 (3) (2009) 701–709.
[42] N. Dilokthanakul, P.A.M. Mediano, M. Garnelo, M.C.H. Lee, H. Salimbeni, K. Arulkumaran, M. Shanahan, Deep unsupervised clustering with Gaussian mixture variational autoencoders, 2016, arXiv CoRR abs/1611.02648.
[43] Z.-H. Sun, T.-Y. Zuo, D. Liang, X. Ming, Z. Chen, S. Qiu, GPHC: A heuristic clustering method to customer segmentation, Appl. Soft Comput. 111 (2021) 107677, http://dx.doi.org/10.1016/j.asoc.2021.107677.
[44] Y.-C. Chang, H. Yang, S. Kong, Based on mini batch K-means clustering for customer segmentation in E-commerce, in: 2022 International Conference on Cloud Computing, Big Data and Internet of Things, 3CBIT, 2022, pp. 60–66, http://dx.doi.org/10.1109/3CBIT57391.2022.00021.
[45] C.S. Reddy, N.S.K. Deepak Rao, A. Sisir, V.S. Srinivasa Raju, S.S. Aravinth, A comparative survey on K-means and hierarchical clustering in E-commerce systems, in: 2023 International Conference on Intelligent Data Communication Technologies and Internet of Things, IDCIoT, 2023, pp. 805–811, http://dx.doi.org/10.1109/IDCIoT56793.2023.10053472.
[46] N.D. Sugiharto, D. Elbert, J. Arnold, I.S. Edbert, D. Suhartono, Mall customer clustering using Gaussian mixture model, K-means, and BIRCH algorithm, in: 2023 6th International Conference on Information and Communications Technology, ICOIACT, 2023, pp. 212–217, http://dx.doi.org/10.1109/ICOIACT59844.2023.10455950.
[47] J. John, O. Shobayo, B. Ogunleye, An exploration of clustering algorithms for customer segmentation in the UK retail market, Analytics 2 (2023) 809–823, http://dx.doi.org/10.3390/analytics2040042.
[48] A. Wasilewski, K. Juszczyszyn, V. Suryani, Multi-factor evaluation of clustering methods for e-commerce application, Egypt. Inform. J. 28 (2024) 100562, http://dx.doi.org/10.1016/j.eij.2024.100562.
[49] S. Greco, M. Ehrgott, J. Figueira, Multiple Criteria Decision Analysis: State of the Art Surveys, Springer, 2016, http://dx.doi.org/10.1007/978-1-4939-3094-4.
[50] S. Chakraborty, TOPSIS and modified TOPSIS: A comparative analysis, Decis. Anal. J. 2 (2022) 100021, http://dx.doi.org/10.31224/osf.io/y39j7.
[51] P.C. Fishburn, Letter to the editor—Additive utilities with incomplete product sets: Application to priorities and assignments, Oper. Res. 15 (3) (1967) 537–542, http://dx.doi.org/10.1287/opre.15.3.537.
[52] W.W. Cooper, L.M. Seiford, J. Zhu, Handbook on Data Envelopment Analysis, Springer Science & Business Media, 2011, http://dx.doi.org/10.1007/978-1-4419-6151-8.
[53] A. Alinezhad, J. Khalili, A. Alinezhad, J. Khalili, COPRAS method, in: New Methods and Applications in Multiple Attribute Decision Making, Madm, Springer, 2019, pp. 87–91.
[54] S. Opricovic, G.-H. Tzeng, Extended VIKOR method in comparison with outranking methods, European J. Oper. Res. 178 (2) (2007) 514–529, http://dx.doi.org/10.1016/j.ejor.2006.01.020.
[55] S. French, Reflections on 50 years of MCDM: Issues and future research needs, EURO J. Decis. Process. 11 (2023) 100030, http://dx.doi.org/10.1016/j.ejdp.2023.100030.
[56] H.-M. Lyu, Z.-Y. Yin, A. Zhou, S.-L. Shen, MCDM-based flood risk assessment of metro systems in smart city development: A review, Environ. Impact Assess. Rev. 101 (2023) 107154, http://dx.doi.org/10.1016/j.eiar.2023.107154.
[57] A. Kumar, B. Sah, A.R. Singh, Y. Deng, X. He, P. Kumar, R. Bansal, A review of multi criteria decision making (MCDM) towards sustainable renewable energy development, Renew. Sustain. Energy Rev. 69 (2017) 596–609, http://dx.doi.org/10.1016/j.rser.2016.11.191.
[58] S. Hosouli, N. Gaikwad, S.H. Qamar, J. Gomes, Optimizing photovoltaic thermal (PVT) collector selection: A multi-criteria decision-making (MCDM) approach for renewable energy systems, Heliyon 10 (6) (2024) e27605, http://dx.doi.org/10.1016/j.heliyon.2024.e27605.
[59] I. Emovon, O.S. Oghenenyerovwho, Application of MCDM method in material selection for optimal design: A review, Results Mater. 7 (2020) 100115, http://dx.doi.org/10.1016/j.rinma.2020.100115.
[60] S.H. Mian, E. Abouel Nasr, K. Moiduddin, M. Saleh, M.H. Abidi, H. Alkhalefah, Assessment of consolidative multi-criteria decision making (C-MCDM) algorithms for optimal mapping of polymer materials in additive manufacturing: A case study of orthotic application, Heliyon 10 (10) (2024) e30867, http://dx.doi.org/10.1016/j.heliyon.2024.e30867.
[61] T. Tan, G. Mills, E. Papadonikolaki, Z. Liu, Combining multi-criteria decision making (MCDM) methods with building information modelling (BIM): A review, Autom. Constr. 121 (2021) 103451, http://dx.doi.org/10.1016/j.autcon.2020.103451.
[62] P. Angaitkar, R. Ram Janghel, T. Prasad Sahu, An MCDM approach for reverse vaccinology model to predict bacterial protective antigens, Vaccine (2024) http://dx.doi.org/10.1016/j.vaccine.2024.04.078.
[63] S.M.A. Iqbal, Asaduzzaman, Cache-MCDM: A hybrid caching scheme in mobile named data networks based on multi-criteria decision making, Future Gener. Comput. Syst. 154 (2024) 344–358, http://dx.doi.org/10.1016/j.future.2024.01.016.
[64] E.K. Zavadskas, Z. Turskis, Multiple criteria decision making (MCDM) methods in economics: an overview, Technol. Econ. Dev. Econ. 17 (2) (2011) 397–427.
[65] Q. Ma, H. Li, A decision support system for supplier quality evaluation based on MCDM-aggregation and machine learning, Expert Syst. Appl. 242 (2024) 122746, http://dx.doi.org/10.1016/j.eswa.2023.122746.
[66] G. Kou, Y. Peng, G. Wang, Evaluation of clustering algorithms for financial risk analysis using MCDM methods, Inform. Sci. 275 (2014) 1–12, http://dx.doi.org/10.1016/j.ins.2014.02.137.
[67] S. Barak, T. Mokfi, Evaluation and selection of clustering methods using a hybrid group MCDM, Expert Syst. Appl. 138 (2019) 112817, http://dx.doi.org/10.1016/j.eswa.2019.07.034.
[68] M.A. Boujelben, A unicriterion analysis based on the PROMETHEE principles for multicriteria ordered clustering, Omega 69 (2017) 126–140, http://dx.doi.org/10.1016/j.omega.2016.08.007.
[69] P. Meyer, A.-L. Olteanu, Formalizing and solving the problem of clustering in MCDA, European J. Oper. Res. 227 (3) (2013) 494–502, http://dx.doi.org/10.1016/j.ejor.2013.01.016.
[70] B. Yatsalo, A. Radaev, L. Martínez, From MCDA to fuzzy MCDA: Presumption of model adequacy or is every fuzzification of an mCDA method justified? Inform. Sci. 587 (2022) 371–392, http://dx.doi.org/10.1016/j.ins.2021.12.051.
[71] H. Aydoğan, V. Ozkir, A fermatean fuzzy MCDM method for selection and ranking problems: Case studies, Expert Syst. Appl. 237 (2024) 121628, http://dx.doi.org/10.1016/j.eswa.2023.121628.
[72] R. Alshahrani, M. Yenugula, H. Algethami, F. Alharbi, S. Shubhra Goswami, Q. Noorulhasan Naveed, A. Lasisi, S. Islam, N.A. Khan, S. Zahmatkesh, Establishing the fuzzy integrated hybrid MCDM framework to identify the key barriers to implementing artificial intelligence-enabled sustainable cloud system in an IT industry, Expert Syst. Appl. 238 (2024) 121732, http://dx.doi.org/10.1016/j.eswa.2023.121732.
[73] A. Baczkiewicz, B. Kizielewicz, A. Shekhovtsov, J. Watróbski, J. Wieckowski, W. Salabun, Towards an e-commerce recommendation system based on MCDM methods, in: 2021 International Conference on Decision Aid Sciences and Application, DASA, 2021, pp. 991–996, http://dx.doi.org/10.1109/DASA53625.2021.9682356.
[74] F. Wang, Novel score function and standard coefficient-based single-valued neutrosophic MCDM for live streaming sales, Inform. Sci. 654 (2024) 119836, http://dx.doi.org/10.1016/j.ins.2023.119836.
[75] B.O. Vieira, P. Guarnieri, R. Nofal, B. Nofal, Multi-criteria methods applied in the studies of barriers identified in the implementation of reverse logistics of E-waste: A research agenda, Logistics 4 (2) (2020) http://dx.doi.org/10.3390/logistics4020011.
[76] A. Wasilewski, Customer segmentation in e-commerce: a context-aware quality framework for comparing clustering algorithms, J. Internet Serv. Appl. 15 (1) (2024) 160–178, http://dx.doi.org/10.5753/jisa.2024.3851.
[77] M.E. Celebi, H.A. Kingravi, P.A. Vela, A comparative study of efficient initialization methods for the k-means clustering algorithm, Expert Syst. Appl. 40 (1) (2013) 200–210, http://dx.doi.org/10.1016/j.eswa.2012.07.021.
[78] T.J. Ross, Fuzzy Logic with Engineering Applications, John Wiley & Sons, 2005.
[79] J.C. Dunn, A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters, 1973.
[80] C.-Y. Tsai, C.-C. Chiu, An efficient feature selection approach for clustering: using a Gaussian mixture model of data dissimilarity, in: Computational Science and its Applications–ICCSA 2007: International Conference, Kuala Lumpur, Malaysia, August 26-29, 2007. Proceedings, Part I 7, Springer, 2007, pp. 1107–1118.
[81] O. Cappé, E. Moulines, On-line expectation–maximization algorithm for latent data models, J. R. Stat. Soc. Ser. B Stat. Methodol. 71 (3) (2009) 593–613, http://dx.doi.org/10.1111/j.1467-9868.2009.00698.x.
[82] H.D. Nguyen, F. Forbes, G.J. McLachlan, Mini-batch learning of exponential family finite mixture models, 2019, arXiv:1902.03335. URL https://arxiv.org/abs/1902.03335.
[83] P. Lingras, G. Peters, Rough Clustering, Wiley Interdiscip. Rev.: Data Min. Knowl. Discov. 1 (1) (2011) 64–72, http://dx.doi.org/10.1002/WIDM.16.
[84] T. Zhang, R. Ramakrishnan, M. Livny, BIRCH: an efficient data clustering method for very large databases, in: ACM SIGMOD Conference, 1996.
[85] J. Figueira, S. Greco, M. Ehrgott, Multiple criteria decision analysis: state of the art surveys, 2005.
[86] S. Hajkowicz, A. Higgins, A comparison of multiple criteria analysis techniques for water resource management, European J. Oper. Res. 184 (1) (2008) 255–265.
[87] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in python, J. Mach. Learn. Res. 12 (2011) 2825–2830.
[88] skfuzzy 0.2 docs, documentation of scikit-fuzzy. URL https://pythonhosted.org/scikit-fuzzy/.