


Customer segmentation in e-commerce: a context-aware quality framework for comparing clustering algorithms

Adam Wasilewski   [Wrocław University of Science and Technology | adam.wasilewski@pwr.edu.pl]

 Faculty of Management, Wrocław University of Science and Technology, Wyb. Wyspińskiego 27, Poland.

Received: 15 November 2023 • Accepted: 03 May 2024 • Published: 25 July 2024

Abstract E-commerce platforms are constantly evolving to meet the ever-changing needs and preferences of online shoppers. One of the ways that is gaining popularity and leading to a more personalised and efficient user experience is through the use of clustering techniques. However, the choice between clustering algorithms should be made based on specific business context, project requirements, data characteristics, and computational resources. The purpose of this paper was to present a quality framework that allows the comparison of different clustering approaches, taking into account the business context of the application of the results obtained. The validation of the proposed approach was carried out by comparing three methods - K-means, K-medians, and BIRCH. One possible application of the generated clusters is a platform to support multiple variants of the e-commerce user interface, which requires the selection of an optimal algorithm based on different quality criteria. The contribution of the paper includes the proposal of a framework that takes into account the business context of e-commerce customer clustering and its practical validation. The results obtained confirmed that the clustering techniques analysed can differ significantly when analysing e-commerce customer behaviour data. The quality framework presented in this paper is a flexible approach that can be developed and adapted to the specifics of different e-commerce systems.

Keywords: e-commerce, personalisation, segmentation, quality framework, clustering

1 Introduction

In the rapidly evolving realm of e-commerce, where millions of products and services are just a click away, consumer attention and loyalty is becoming one of the most important goals to achieve. The era of *one-size-fits-all* user interfaces (UI) and generic shopping experiences is giving way to a new paradigm called **personalisation**. E-commerce personalisation is more than a marketing approach. It is a dynamic and data-driven strategy that is revolutionising the way online retailers interact with their customers and refers to the selection of content based on customer characteristics to improve business results for an e-commerce platform [Aksoy *et al.*, 2023]. The origins of this concept can be traced back to the end of the 20th century, when online shops took their first steps in presenting product recommendations to customers. The simple algorithms used initially were replaced over time by more complex solutions based on machine learning and artificial intelligence methods. Technological advances in data collection and processing have also been significant, making it possible to collect and analyse huge amounts of customer data. This data included not only purchase history, but also click patterns, search queries and demographic information. The ability to harness this data has played a key role in e-commerce personalisation efforts. Of particular importance has been the growth in the use of advanced algorithms to interpret and respond to user behaviour. Such solutions make it possible to predict customer preferences and make real-time decisions about content and UI adjustments. Moreover, different aspects of personalisation can induce cognitive and hedonic user experiences when interacting with web-

sites, which in turn generate satisfaction and influence the user's decision to revisit the personalised website [Desaid, 2019].

One of the oldest techniques for individualising user targeting in e-commerce is segmentation, which involves dividing a visitor base into distinct groups based on different characteristics and behaviours. This allows companies to better understand their customers and tailor marketing strategies, product offerings and experiences to each segment's unique needs [Camilleri, 2017]. The segmentation is based on the information collected, which may come directly from the users or be collected indirectly as a result of them using the webshop pages. Depending on the data, segmentation can be based on demographics (age, gender, income, education), geography (location, climate), behaviour (purchase history and frequency, website decisions, loyalty), technology (devices and software used), psychology (lifestyles, personalities) or channel (online, mobile, retail, business-to-business, business-to-consumer) [Dolnicar *et al.*, 2018]. Customer segmentation using clustering - a data-driven approach to grouping customers into clusters based on similarity of attributes or behaviour - has also grown in popularity over the past few years [Gomes and Meisen, 2023]. When using clustering for customer segmentation, it's important to assess the quality of the resulting clusters, validate them against business objectives, and continually monitor and update the segments as customer behaviour evolves [Punhani *et al.*, 2021].

Existing personalisation solutions in e-commerce focus on product recommendations [Xiao and Benbasat, 2007], which influence the design of the user interface, but only to a very limited extent [Kopel *et al.*, 2013]. Fully adapting the de-

sign of the layout to the characteristics and behaviours of customers is a much more difficult process and therefore much less widely used [Wasilewski and Przyborowski, 2023]. For this reason, there is a lack of in-depth analysis of clustering methods that are worth implementing for the delivery mechanism of personalised web shop designs. However, in order to identify the optimal clustering method under specific business conditions, it is necessary to identify the factors that determine the results obtained and may represent potential limitations to their practical application [Faraone *et al.*, 2012]. Customer segmentation for the personalisation of the e-commerce user interface should take into account three main aspects to assess the suitability of the algorithm for specific business needs. They include: the resources required, the context-free quality of the clustering and the convergence of the characteristics of the clusters obtained with the specificity of serving a dedicated e-commerce user interface. It should be noted that the choice of evaluation criteria may vary depending on the specific application of clustering [Gomes and Meisen, 2023]. As the choice of segmentation method is critical to the business effectiveness of each type of solution, a multi-dimensional review of the similarities and differences between possible approaches can significantly influence the perception of the quality of the overall comprehensive solution. The research discussed in this paper has focused on three algorithms, K-means, K-medians, and BIRCH, on the basis of which a practical application of the proposed quality framework for comparing clustering methods is presented. The choice of these methods was not arbitrary. Two of them (K-means and K-medians) are classified as partitioning methods, and the similarity of their names can sometimes lead to the conclusion that the results obtained from their application will also be similar [Gomes and Meisen, 2023]. Both are also commonly used to segment e-commerce customers. Therefore, their comparison will identify aspects where the chosen algorithms are indeed similar and where they differ, while validating the proposed approach itself. In addition, a third clustering technique, BIRCH, was included in the study. It is categorised as a hierarchical method and its use allowed a broader verification of the proposed quality framework. The main contributions of this paper include (1) the proposal of a quality framework (taking into account the proposed context-aware metrics) that allows the comparison of clustering methods in a specific business application, (2) the evaluation of metrics that can be used to assess the quality of clustering results, and (3) the analysis of selected algorithms based on experimental studies and aggregation of partial results. Two approaches to multi-criteria decision making have been proposed and used - simple ranking analysis and TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) [Hwang and Yoon, 1981]. The paper is structured as follows. Section 2 provides an overview of the research work related to clustering applications in e-commerce and a description of the three algorithms selected for in-depth study. Section 3 presents an extended description of the problem, its business relevance, and the proposed quality framework. Section 4 describes the experimental study, its results and lessons learned. Section 5 concludes the paper and gives directions for further research.

2 Literature review

Clustering is proving to be a valuable tool in the e-commerce industry, providing businesses with insights and opportunities to improve the customer experience, optimise operations and increase profitability. By organising data into meaningful clusters, e-commerce companies can make more informed decisions, ultimately contributing to their success in a highly competitive marketplace. Clustering methods are used in various areas of e-commerce such as customer segmentation, product recommendations, inventory management, fraud detection, market trend analysis, personalisation and more. This section provides an overview of selected work related to clustering applications in e-commerce and gives an overview of the algorithms that have been used in experimental studies.

The most common application of clustering algorithms in e-commerce is their use for customer segmentation. In contrast to traditional segmentation based on decision rules (*rule-based segmentation*), it can be referred to as *clustering-based segmentation* [Nurma Sari *et al.*, 2016]. This approach is useful when hidden patterns need to be discovered or when data segmentation criteria are not well established. It can be particularly suitable for exploratory analysis, or when looking beyond pre-defined rules. Different approaches to segmentation using clustering methods can be found in the literature. These include the analysis of relatively simple data sets, such as purchase data [Wu and Chou, 2011], as well as extensive behavioural data covering the entire customer activity of an online shop [Su and Chen, 2015]. When considering the specifics of serving dedicated user interfaces, the latter application is particularly noteworthy. In this case, the analysis should also cover the entire history of the customer's activity, sometimes referred to as the *clickstream*. This can be most simply defined as a sequence of clicks made by a user [Wang *et al.*, 2017], but in practice it can be interpreted differently. In some applications, it may be limited to logging the pages users visited, the time spent on each page, how they arrived at the site and where they went next [Albert *et al.*, 2010]. More broadly, this may include recording all activities performed by a visitor, such as links, buttons and images clicked, pages viewed, forms submitted and other actions taken on the website or application. It may also include information about the user's device, location and referral sources [Koehn *et al.*, 2020]. It should be noted that the effective operation of personalisation technology in e-commerce today depends on the usefulness of personalised recommendations and on consumers' privacy concerns or preferences in trading personal information [Song *et al.*, 2021]. Customers are increasingly aware of the rules governing the collection of behavioural data. In addition, legislation (particularly in the European Union) and market trends (minimising the importance of third party cookies) mean that the collection of behavioural data, which is critical for advanced customer segmentation, requires additional effort. In some applications, the privacy issue can be circumvented by using information that customers voluntarily provide in the system, such as adding online reviews [Wu and Liu, 2020]. The use of generative artificial intelligence in personalisation should also not be overlooked, as it has both potential and

risks [Ooi *et al.*, 2023]. The problem of segmentation using clustering algorithms is also the subject of research in logistics, including e-commerce logistics [Wasilewski, 2019]. It can be used to analyse historical data such as order history, product attributes, delivery routes, etc., and the result can be the optimisation of various functional aspects. Examples of these include the creation of efficient distribution routes, taking into account delivery and customer satisfaction [Zheng *et al.*, 2023], solving a two-stage location routing problem in last-mile delivery [Amini and Haughton, 2023], improving cross-docking functions [Amna Altaf and Lecoutre, 2023], and dynamic container drayage booking and routing decision support [Chen *et al.*, 2023]. It should be noted that while some publications approach the topic more generally, focusing on the development of frameworks using clustering in areas such as returns handling [Nanayakkara *et al.*, 2022] or supply chain management in general [Mashalah *et al.*, 2022], there are also descriptions of clustering applications in logistics in the direct context of e-commerce [Hjort *et al.*, 2016]. Among the clustering algorithms used in practice to segment e-commerce customers, *K*-... (e.g., *K*-means, *K*-medoids, *K*-medians) methods can be distinguished by popularity [Gomes and Meisen, 2023]. In addition to its frequent use in typical tasks leading to customer segmentation [Guo and Altrjman, 2022; Li *et al.*, 2022; Solichin and Wibowo, 2022; Zhao *et al.*, 2021], *K*-means approach can be employed in more detailed applications. It has been used to process data containing relationships between three sets of data: event type, products and categories, and has allowed some limitations to be identified [Tabianan *et al.*, 2022]. Another study showed that a *K*-means based solution allows consumers to dynamically adjust their preferences and combine information from different sources to identify products that are overpriced or otherwise dominated by competing alternatives [Papamichail and Papamichail, 2007]. It is also possible to use this clustering method effectively for the analysis of social e-commerce data, for which traditional user classification methods are not suitable [Cui *et al.*, 2021]. This hypothesis is in line with the findings of other studies [Shen, 2023]. *K*-means is sometimes used to analyse customer satisfaction in e-commerce [Zare and Emadi, 2020], which is particularly important as the trends and developments move towards a customer-centric market [Meena *et al.*, 2023]. This approach can also be used when segmenting e-commerce customers by their communication channel (desktop versus mobile) [Rajput and Singh, 2023]. Another interesting and promising area of research where this algorithm can be applied is in the area of consumption behaviour of e-commerce customers. This feature engineering approach has led to the clustering of customers into four categories ('iron powder customers', 'general customers', 'develop customers' and 'zombie customers') and has made it possible to assess their relevance to the company [Zhang *et al.*, 2022]. This algorithm can also be used with classic customer segmentation approaches such as the RFM (Recency, Frequency, Monetary Value) model to recommend relevant marketing strategies for the e-commerce industry [Ma, 2022]. However, *K*-means applications are not limited to analysing consumer behaviour, but can also be useful in segmenting business partners working together in a B2B model [Punhani *et al.*, 2020].

Another popular centre-based clustering algorithm is *K*-medians. The main difference is that the centre is determined by the median, unlike *K*-means where the centre is determined by the mean. *K*-medians is also sometimes used in e-commerce applications. It can be used to segment customers based on transactional data, with promising results in terms of the proportion of cluster sizes generated [Maulana *et al.*, 2023]. An example application is also the streaming fraud detection solution, where the approach is used for a cache of frequently used subgraphs [Nguyen *et al.*, 2023]. Another example of the use of *K*-medians clustering for predictive rating is the proposal for an online book recommendation system [Okon *et al.*, 2018]. It is also possible to find applications of *K*-means and *K*-medians within the same research, for example in the clothing industry [Tsao *et al.*, 2023] and B2B (Business 2 Business) customer service [Tsao *et al.*, 2022]. However, when choosing an algorithm for e-commerce applications, it is important to remember that *K*-medians may be less efficient than *K*-means in terms of convergence speed, especially on large datasets [Han *et al.*, 2024].

The BIRCH (Balanced Iterative Reduction and Clustering using Hierarchies) algorithm is sometimes used for customer segmentation [Fontanini and Abreu, 2018] and it is most often one of the techniques compared in specific applications. An interesting perspective on the comparison of different clustering methods in e-commerce is presented by John *et al.* [2023]. Based on methods: *K*-means, the Gaussian Mixture Model (GMM), Density-Based Spatial Clustering of Applications with Noise (DBSCAN), Agglomerative Clustering and BIRCH, it was found that the best results (as measured by the Silhouette Score value) were given by GMM. However, it is worth noting that the conclusions were based on only one indicator of context-free clustering quality, ignoring other indicators in this group and not taking into account computational complexity and business context suitability. Slightly different results were obtained when comparing the BIRCH, Agglomerative Clustering, *K*-Means and DBSCAN algorithms for customer segmentation [Sahinbas and Catak, 2022]. The results of these studies showed that all approaches gave almost the same clustering results, but DBSCAN proved to be the best in terms of Silhouette value. The *K*-means, BIRCH and DBSCAN algorithms were also compared for recommending films based on customer feedback [Nawara and Kashef, 2021]. In this case, the factors considered were MAE (Mean Absolute Error), RMSE (Root Mean Square Error) and computed time, which is an attempt at multi-criteria analysis. Multi-criteria clustering quality comparisons were also performed for DBSCAN, *K*-means, Mini Batch *K*-means and Mean Shift [Hicham and Karim, 2022]. Adjust Rand Index (ARI), Normalised Mutual Information (NMI), Dunn's Index (DI) and Silhouette Score were used as criteria for differentiation, but all refer to context-free quality.

An analysis of the literature shows that comparative studies of clustering methods used in e-commerce focus on context-free metrics that apply regardless of the business use of clusters. However, when selecting an algorithm for customer segmentation, other aspects need to be taken into account, both in terms of business context and computational complexity (which has a direct impact on the computing resources

Table 1. Key differences between K-means, K-medians and BIRCH
[Han *et al.*, 2011; Dasgupta *et al.*, 2020]

Differentiating factor	K-means	K-medians	BIRCH
objective	minimise the sum of the squared distances (Euclidean distances) between data points and the centroid of their assigned cluster	minimise the sum of the absolute deviations of data points from their respective cluster medians	to build a compact summary of the dataset by recursively clustering data points into smaller and denser clusters while maintaining the global structure of the data
representation	each cluster is represented by the mean (centroid) of the data points belonging to that cluster	each cluster is represented by the median of the data points belonging to that cluster	clusters are represented using a hierarchical tree structure, and each non-leaf node in the tree represents a cluster
outliers	sensitive to outliers as mean is affected by extremes, resulting in biased clusters	more robust to outliers than k-means since it minimizes the sum of absolute deviations, which is less sensitive to extreme values	relatively robust to outliers due to its ability to incrementally merge clusters and construct a hierarchical structure
comprehensibility	lack of a specific, real reference point in the cluster, which makes it difficult to interpret the clusters	depends on the representation of centroids, cluster size and shape	influenced by its hierarchical structure

used). For this reason, the lack of comprehensive analyses of the various factors influencing the final choice of clustering method can be considered a research gap. Accordingly, this paper addresses this issue by defining a quality framework based on different decision criteria and by applying it in an experimental study comparing three clustering algorithms.

3 Problem statement

3.1 The importance of context-aware clustering

Clustering is an important machine learning technique used to group data points into clusters with similar characteristics. Among the various clustering algorithms, the K-means, the K-medians, and BIRCH approaches are the commonly used. The first two centroid-based approaches have much in common, but there are also important differences between them [Dasgupta *et al.*, 2020; Sihombing, 2021]. The third technique belongs to the group of hierarchical methods [Lorbeer *et al.*, 2017] and, for this reason, significant differences in results should be discernible in experimental studies. The main differences between described algorithms are shown in the Table 1. Such a selection of clustering methods will allow the verification of the quality framework for both similar and significantly different techniques.

An additional motivation for the study is the importance of the datasets characteristics and the context in which clustering is applied. None of the existing comparisons of the three methods explicitly address the clustering of e-commerce customers on the basis of their complex behaviour. Meanwhile, this issue is important from the perspective of personalisation, which has become one of the major trends in recent

years, setting new standards in customer communication. One way to personalise is to offer multi-variant user interfaces in the web shop. This is a move away from the one-size-fits-all approach to tailoring the content and design offered to the behaviour, choices or preferences of the visitor. A comprehensive solution for the design and delivery of dedicated UI variants (Figure 1) requires the implementation of many functions, and one of the key elements is the segmentation of customers according to their characteristic e-commerce behaviour. The overall framework and potential business benefits have been outlined in other studies [Wasilewski, 2024], but the issue of selecting a clustering method is an additional challenge. In order to group the users of an online shop so that the resulting clusters of customers can be applied and served with a dedicated user interface, a decision has to be made about the clustering method. On the one hand, it is a choice between different clustering approaches and, on the other, the choice of a specific algorithm and its parameters. An analysis of the prevalence of practical applications of the various methods and comparative studies allows a preliminary selection of potentially suitable solutions, but does not always provide a clear answer which is optimal for a specific business application. This may be due to differences in the business context and structure of the data analysed, as well as in the research approach itself.

Taking these factors into account, an attempt was made to compare the results of implementing three clustering methods - K-means, K-medians, and BIRCH - for generating clusters of e-commerce customers to be served with dedicated user interface variants. The analysis covered three aspects that can significantly influence the decision to choose a particular algorithm - the consumption of computing resources, the quality of clustering and the degree to which the requirements arising from the business context are met.

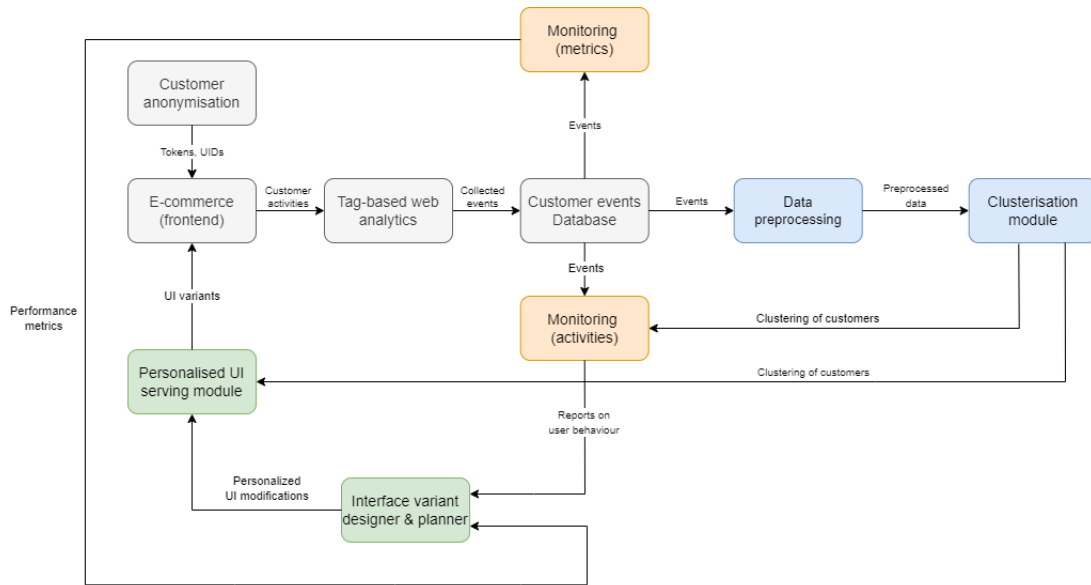


Figure 1. Functional framework for multi-variant e-commerce user interfaces [Wasilewski, 2024].

Clustering is generally a time-consuming and computationally intensive operation. This is one of the main limitations of the technique. Computational complexity should always be taken into account when choosing an algorithm, but even between algorithms of similar complexity there may be differences due to the specifics of the data being processed, the implementation of the algorithm, or the initial parameters set. In the case of the K-means and K-medians algorithms, theoretically there should be no significant differences in clustering time and resource consumption, but empirical verification allows this hypothesis to be tested. The BIRCH method, on the other hand, is expected to use resources differently from the two centric-based techniques.

Cluster quality measures are used to assess the quality of clusters produced by various clustering algorithms. By measuring factors such as cluster separation, cohesion and compactness, these metrics help to quantify how well a clustering solution performs. Commonly used clustering quality indicators are Silhouette Score, which measures the separation between clusters, Davies-Bouldin Index, which quantifies the average similarity between each cluster and its most similar cluster, and Calinski-Harabasz Index, which measures the ratio of between-cluster variance to within-cluster variance. A comparison of the values of these indices for the algorithms tested allows a context-free evaluation of the clustering results. The results obtained can support the recommendation of one of the methods, but also show the relationships between the main initiating parameter (the number of clusters) and the evaluation of the clustering quality.

The business context of the clustering application is an important but often underestimated factor in the choice of approach to be implemented. In this case, there are no common metrics that can be used explicitly. In order to compare methods, it is necessary to look at the specifics of the business and the way clusters are applied. From an economic point of view, the key limitation is the number of clusters. Given that there should be a dedicated UI variant for each cluster, it should be assumed that there will not be too many, as each additional one means additional design and implementation

costs. The next requirements are indirectly related. Firstly, the resulting clusters should contain a minimum reasonable number of webshop customers (defined as a relative or absolute value), since given the percentage of returning users, with clusters of small size the dedicated UI variant will be used sporadically. Secondly, the clusters should be as similar in size as possible, so that all the prepared UI variants are served with similar frequency. This requirement is particularly important if a self-adaptation mechanism for the UI is to be implemented, because the speed of obtaining feedback from customer behaviour and decisions will determine the tuning time of the system. As there are no studies that address the suitability of the clustering results for application in the described business context and with the resulting structure of the customer behaviour datasets, for the evaluation of all methods, dedicated indicators were defined and the conclusions were based on them. This set of decision factors is one of the aspects of the novelty of the research that is described in this paper.

The final decision on the choice of a particular clustering method should be the result of an analysis of all the aspects mentioned. Some of the requirements should be hard constraints (e.g. clustering time, resources consumed, number of clusters, minimum size of the smallest cluster), disqualifying combinations of methods and starting parameters that do not meet them. In turn, the second set of requirements, applied to the other clusterisation options, should allow a comparative analysis of the obtained values of the selected indicators, with defined priorities.

In order to systematise a comparative study of the three exemplary clustering algorithms, the quality framework has been proposed. It takes into account the aspects described above, which affect the overall evaluation of the analysed options. An experimental study was carried out to validate this proposal and verify its practical suitability, taking into account the quality metrics included in the framework.

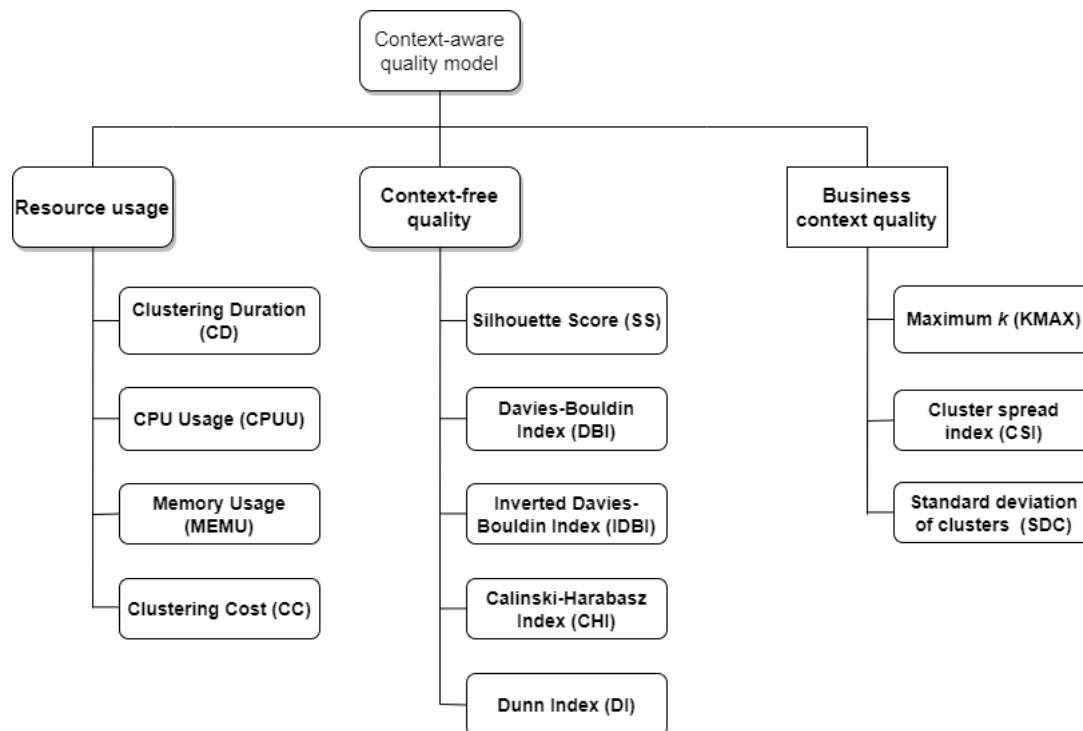


Figure 2. Context-aware quality framework for clustering algorithms.

3.2 Description of the method

The core of the proposed approach for comparing clustering algorithms for e-commerce customer segmentation is the quality framework that contains a set of characteristics and metrics to quantify them (Figure 2). This concept is consistent with typical software quality models such as ISO/IEC 9126 [Al-Kilidar *et al.*, 2005] or ISO/IEC 25010 [Estdale and Georgiadou, 2018].

Resource usage

Three main characteristics are included in the framework, which group together metrics that describe the partial (related to a specific characteristic) quality of clustering.

Characteristics refer to the resource consumption during the clusterisation. Computational complexity is an intrinsic property of any algorithm and is one of the key features considered when selecting a clustering method. It translates into the size of the datasets that can be processed, so it cannot be neglected in the proposed quality framework.

It is important to note the practical aspect of this characteristic. The possible use of a clustering algorithm takes place in a specific computing environment, which can be dimensioned by the available RAM, number of processes and threads, and disk resources.

For typical servers, these resources are usually difficult to increase when clustering becomes too computationally demanding. It requires changing components (for on-premises solutions) or changing vendor contract terms (for virtual servers), which is neither logical nor simple if clustering is infrequent and the additional resources would not be used most of the time. The solution may be to use scaled cloud resources, but even these have their limitations. The main one is the cost of the resources used, and some algorithms can consume huge amounts of resources when working on large datasets. This means that the costs associated with clustering

can rise rapidly, especially if no quota limits are set.

In terms of resource consumption, algorithms that produce results quickly, using less memory or processor power, should be considered better. Therefore, some metrics have been proposed to quantify this clustering quality attribute, allowing direct comparisons between algorithms (Table 2).

The choice of metrics for comparing clustering algorithms should depend on the availability of resources and the resource provisioning model. In the case of servers with fixed available resources, the *CD* metric should be critical, as it allows the selection of a less time-consuming and therefore less resource-intensive approach. The *CPUU* and *MEMU* metrics, on the other hand, are important for algorithms with high computational complexity or large amounts of processed data. The use of all available resources during clustering (indicator values close to 100%) means that there may be problems with the performance or even reliability of the solution. When comparing different algorithms, choose those that do not require the use of all available resources. *CC* is important when using resources where the cost is determined by usage, such as cloud servers. In such cases algorithms with lower clustering costs should be preferred.

Context-free quality

Context-free quality indicators for clustering methods are a well-studied area for comparative analysis of different algorithms. Some of the most common examples used in practice are included in the proposed quality framework (Table 3), but this list is not closed and can be freely extended.

Two aspects of the use of classical context-free clustering quality metrics in the described approach are worth noting. First, the *IDB* metric has been introduced, which is the inverse of the *DB* metric. The purpose of this operation is to make the interpretation of the results obtained more consistent. For metrics other than *DB*, a higher value implied

Table 2. Resource usage metrics

Metric	Description	Interpretation	Comments
Clustering Duration (<i>CD</i>)	Time required for clusterisation, excluding pre-processing time.	The shorter the time needed for clusterisation the better.	in seconds
CPU Usage (<i>CPUU</i>)	Maximum consumption of available processor computing resources during clustering.	The lower the maximum processor resources the better.	percentage
Memory Usage (<i>MEMU</i>) e	Maximum consumption of available memory resources during clustering.	The lower the maximum memory consumption the better.	percentage
Clustering Cost (<i>CC</i>)	Cost of performing clustering, paid to the server service provider.	The lower the clustering costs the better.	in currency

better clustering, and with the *IDB* metric it is possible to take advantage of the *DB* metric while interpreting the results in a consistent manner.

The second issue concerns the comparability of the values of the quality indicators. With the exception of the *SS* metric, which always gives results between -1 and 1, the values of the other metrics are generally unpredictable (although always positive, which is a simplification). Therefore, they can be standardised by dividing the calculated value of a metric by the maximum value of that metric. An example of the application of this approach is the selection of the optimal number of clusters (the value of the parameter *k*). Once the clustering simulations have been carried out for different *k*, the results can be standardised to bring them into the same range (e.g. from 0 to 1 or from 0 to 100). Thus processed, the values of the different metrics of context-free clustering quality can be aggregated to a single indicator, e.g. by applying weights.

Business context quality

Business context metrics can be used to assess the usefulness of clustering results in practical applications. In the case of e-commerce customer segmentation, several key characteristics can be identified that may influence the perception of this aspect of clustering quality (Table 4).

$$CSI = \frac{CS_{max} - CS_{min}}{CM_{max}} \quad (1)$$

where: CS_{max} - size of the most numerous cluster, CS_{min} - size of the least numerous cluster

The proposed metrics aim to quantify aspects of segmentation that allow the generated customer groups to be used to design dedicated and tailored user interface variants for them. For this reason, the main constraints related to the number and size distribution of the outcome clusters.

Serving multi-variant UIs is based on the concept of presenting specific groups of clients with versions of the layout that are specific to them. The modifications prepared for them should be based on the characteristics that differentiate the behaviour of users from various clusters and the similarities between clients placed in the same cluster.

In this approach, the number of clusters is therefore crucial. It should represent a compromise between the desire for the

best possible personalisation (which points towards as many clusters as possible, up to so-called hyper-personalisation, which can be interpreted as one-element clusters and serving each customer a unique UI variant) and economic rationality (which points towards several clusters, as each variant represents an additional cost). From a business point of view, it can be assumed that it makes sense to prepare a dedicated UI variant if the number of customers (cluster size) that can be served by it is sufficiently large. The acceptance threshold can be defined in absolute terms (number of customers in the cluster) or relative terms (percentage of the total customer population) and depends on individual business preferences. In general, it can be assumed that the higher the threshold is adopted, the smaller the number of outcome clusters will be. To measure this aspect of clustering quality, a metric *KMAX* has been proposed, which represents the highest number of clusters *k* where the smallest cluster has a size above the assumed threshold. When comparing different clustering algorithms, the one with the higher *KMAX* should be considered the better one, as it offers the possibility to serve more dedicated UI variants.

The next two proposed metrics relate to the distribution of cluster sizes. From a business perspective, one should aim for a situation where the resulting clusters are of similar size. This issue can be considered in two ways - the difference between the size of the largest and smallest cluster, and the standard deviation (or variance) of the size of the clusters. The first is addressed by the introduced metric *CSI*, which is calculated from the size of the smallest and largest cluster. The second is the standard deviation (*SDC*). As mentioned above, the size of the clusters depends on the number of clusters adopted, so the use of these two metrics to compare clustering algorithms should be for the same value of *k*.

The metrics presented here allow an assessment of the business context quality of clustering, but obviously do not exhaust all aspects related to customer segmentation in e-commerce. The proposed framework can be extended in the future to include additional metrics that take into account other needs arising from other applications of clustering algorithms. The initial selection of metrics was linked to the business needs of a specific practical application - serving dedicated UI variants to e-commerce shop customers.

A separate issue is the aggregation of the sub-scores and the

Table 3. Context-free quality metrics

Metric	Description	Interpretation	Comments
Silhouette Score (<i>SS</i>) [Rousseeuw, 1987]	A measure of the degree of cohesion and separation of data points within clusters.	A negative score indicates that the datapoint has probably been misclassified. Scores near 0 indicate that the data points are at or near the dividing line. The more the score is greater than 0, the better grouped the data point is, and its distance to the points of its assigned cluster is less than its distance to the points of the nearest neighbouring cluster.	the score ranges from -1 to 1
Calinski-Harabasz Index (<i>CHI</i>) [Calinski and Harabasz, 1974]	A measure of how similar an object is to the cluster it belongs to, compared to other clusters.	A higher value is indicative of better clustering, as this means that the data points are better distributed across a cluster than within a cluster.	
Dunn Index (<i>DI</i>) [Dunn, 1973]	A measure of the compactness of clusters (intra-cluster similarity) and the separation between clusters (inter-cluster dissimilarity).	A higher value indicates better clustering.	
Davies-Bouldin Index (<i>DBI</i>) [Davies and Bouldin, 1979]	A measure of the compactness and separation of clusters in a clustering result.	The lower the index value, the better the clustering performance, as inter-cluster separation increases and intra-cluster variation decreases.	
Inverted Davies-Bouldin Index (<i>IDBI</i>)	The inverse of the <i>DBI</i> indicator, for easier interpretation of the results.	The higher the index, the better clustering.	

resulting decision on the choice of clustering method. There are a number of approaches to making a decision based on multi-criteria analysis. For the purposes of this study, two of these were selected and applied - simple ranking and the TOPSIS method. The results were used to evaluate the different clustering methods and to interpret the results. In addition, the use of aggregation methods allows the sensitivity of the proposed quality framework to be verified. Based on the indicators describing the clustering, it is possible to see how the final results are influenced by individual quality characteristics. This part of the study was limited to the TOPSIS method and examined what the ranking of clustering methods would be if only context-free metrics were used (which is the most popular approach in the literature) and what the changes would be if performance metrics were also included. Finally, these results were compared with a recommendation based on the whole framework (including context-based metrics) and conclusions were drawn.

4 Experimental study

The purpose of the experimental study was to validate the proposed clustering quality framework and to compare the three algorithms in terms of their suitability for generating groups of customers to be served a dedicated UI variant. Experimental studies were carried out on separate datasets

containing the customer activity history of two online shops. The data collected included all actions taken by users and was therefore in the form of a clickstream and the collection process used a combination of two tools - Google Tag Manager (GTM) and a customised version of the Matomo system. A total of 512,355 customer sessions from the first store (146 days in a medium-sized e-shop, Dataset A) and 532,576 customer sessions from the second store (9 days in a large e-shop, Dataset B) were collected and used as a learning dataset for the clustering methods studied. Clustering was carried out using a server with the following hardware parameters:

- CPU: AMD Ryzen 5 3600 6-Core Processor,
- RAM: 128 GB DDR4,
- Disk: 2 x 1024 GB NVMe SSD,
- Connection: 1 GBit/s port.

The servers were not loaded with other services at the time of the study.

4.1 Methodology

The first part of the research involved the collection of data on customer behaviour and decisions. These were organised in a structured form and included the following information:

- session - session's ID,
- userId - customer's unique ID,

Table 4. Business context quality metrics

Metric	Description	Interpretation	Comments
Maximum ($KMAX$)	k The maximum number of clusters (k) at which the size of the smallest cluster exceeds an accepted threshold value (thl).	The higher the value, the better.	The threshold value thl can be set absolutely (as the number of users in the cluster) or relatively (as a percentage of the population of clustered users).
Cluster spread index (CSI)	A measure of the spread of cluster sizes.	The lower the value, the better.	Equation 1
Standard deviation of clusters (SDC)	Standard deviation of cluster sizes, which measures the variation in the number of users in the clusters.	The lower the value, the better.	

- type - type of user activity, such as event, listing, product, homepage, checkout, blog or other,
- category - optional activity category for events, such as product page, cart, search, purchase, etc.,
- action - optional activity action for event, such as mini-cart, select size, select color, thumbnail click, click accordion, image click, etc.,
- name - optional additional description of the action, such as open, close, size value, color value, etc.,
- url - web address of the page where the activity took place,
- time - date and hour of the action.

Despite the simplicity of the data structure, it was possible to record all the actions taken by users of the web shop, from choices that did not lead to a page change (e.g. selection of a filter value, selection of a product size) to the use of the search engine and page transition.

The second part of the research, directly related to clustering, was preceded by pre-processing. This operation, which is the preparation of the data for the actual analysis, involved cleaning and formatting the data to provide the clustering algorithms with a consistent set of learning data. This step involved removing duplicates, handling missing values and normalising the data. In addition, dimension reduction was performed by omitting features irrelevant to clustering. The time required for pre-processing was not included in the comparative analysis of the methods tested, as it is independent of the algorithm chosen.

In order to compare the clustering results obtained after applying the analysed algorithms, an experimental study was carried out by clustering the two data sets with a varying parameter k , ranging from 3 to 10. After each clustering, a set of information was collected to compare the results.

A comparison of methods was carried out for each of the characteristics included in the proposed quality framework - resource usage, context-free and business context. Three metrics were included in each of the outcome aspects evaluated, giving a total of nine dimensions (CD , $CPUU$, $MEMU$ and SS , CHI , IDB and $KMAX$, CSI , SDC) that differentiate the clustering algorithms compared. The algorithms were compared separately for each metric. Where necessary, additional assumptions were made about the number of re-

sulting clusters.

In addition, because the research was conducted on two different learning datasets, the results can be treated independently. This means that they can be considered as a double check when drawing conclusions and planning further research. Furthermore, with this approach, it was possible to observe the impact of the learning data on the clustering algorithms. Although the structure of the data was the same and the number of user sessions analysed was very similar, the values obtained for some quality indicators differed significantly.

4.2 Results

Resource usage

The calculated values of quality metrics related to resource consumption during clustering are shown in Table 5. The values of the CD indicator are given in seconds and include the clustering time and the generation of the final reports (including visualisation of the results). When analysing the results, it can be seen that the values of this metric were highest for K-medians clustering and lowest for K-means, regardless of the number of outcome clusters. The difference between K-means and K-medians was about 85% and was similar for both datasets. The clustering time for the BIRCH method was in between the two and was about 30% higher than for K-means. This leads to the conclusion that the K-means algorithm is superior in terms of clustering time and should be recommended for use if only the CD indicator criterion were used.

The results of the study also allow for two additional observations. First, the values of the CD index varied markedly with the number of clusters used (k), but these differences do not form a pattern. The differences between the shortest and the longest clustering time within the [algorithm - dataset] combination ranged from 8.2% (BIRCH - Dataset A) to 22.1% (K-medians - Dataset A). For both datasets, the smallest differences in clustering time for different values of k were observed for the BIRCH method and the largest for the K-medians method. It is worth noting that these differences were not of great significance during the course of the research, but the CD metric may be important when comparing algorithms that need much more time for clustering.

Table 5. Resource usage comparison

DATASET A									
<i>k</i>	<i>CD</i>			<i>CPUU</i>			<i>MEMU</i>		
	K-means	K-medians	BIRCH	K-means	K-medians	BIRCH	K-means	K-medians	BIRCH
3	597	1136	780	97.35%	97.27%	90.93%	37.71%	38.95%	44.90%
4	617	1136	801	97.60%	97.26%	94.12%	38.55%	38.66%	44.88%
5	606	1238	770	83.93%	82.82%	89.44%	38.05%	38.67%	45.71%
6	587	1044	769	93.92%	77.92%	94.73%	38.44%	38.40%	45.67%
7	565	1127	821	83.98%	83.04%	99.81%	38.15%	38.93%	45.65%
8	607	1066	832	97.41%	90.10%	97.42%	38.26%	38.72%	45.63%
9	596	1014	821	97.34%	88.77%	90.71%	38.39%	38.50%	46.51%
10	567	1045	791	82.77%	87.70%	92.77%	37.96%	38.32%	45.65%
avr.	592.75	1100.75	798.13	91.79%	88.11%	93.74%	38.19%	38.65%	45.45%
std.dev.	18.75	72.78	24.56	6.92%	6.86%	3.55%	0.28%	0.23%	0.35%
DATASET B									
3	1674	3458	2339	99.96%	99.96%	99.91%	43.13%	41.71%	57.81%
4	1768	3343	2185	99.93%	99.94%	99.76%	42.53%	41.60%	57.69%
5	1758	3552	2185	99.92%	99.85%	99.96%	42.56%	41.50%	58.71%
6	1677	3144	2340	99.93%	99.98%	99.88%	42.42%	42.60%	58.59%
7	1922	3243	2246	99.93%	99.91%	99.89%	43.14%	42.56%	58.55%
8	1943	3008	2390	99.96%	99.95%	99.90%	42.31%	42.53%	58.56%
9	1715	3307	2297	99.92%	99.96%	99.93%	42.21%	42.55%	58.57%
10	1788	3371	2503	99.96%	99.87%	99.93%	41.94%	42.56%	58.60%
avr.	1780.63	3303.25	2310.63	99.94%	99.93%	99.89%	42.53%	42.20%	58.39%
std.dev.	102.46	172.45	107.51	0.02%	0.05%	0.06%	0.42%	0.50%	0.40%

The conclusion of this observation is to confirm the need to calculate the *CD* index for different values of *k* to avoid the risk of falsifying the result with a single measurement.

The second interesting observation was the significant differences in the *CD* values for the two datasets. It might seem that with similar sized datasets and the same learning data structure, the differences in clustering time should not be significant. However, for both algorithms, it took 2.8-2.9 times longer to cluster Dataset B than it did to cluster Dataset A. The explanation for this situation lies in the specifics of the online shops where customer behaviour data was collected for both datasets. In case A, it took 146 days to collect more than 500,000 user sessions, during which time 110797 unique customers appeared. In case B, on the other hand, it took nine days to collect a similar amount of learning data, and in that time 212431 users were identified, which is about 92% more than in A. The number of objects (customers) to be clustered is therefore the main reason for the increase in clustering duration, with the increase in clustering duration being disproportionately greater than the increase in the number of users. This means that when estimating the resources required to implement clustering, the focus should be on the number of webshop customers and their retention, and the size of the dataset, however important, should be secondary. The values of the *CPUU* and *MEMU* indicators are in percentage and describe the maximum resource usage during clustering. Both metrics were sampled every 60 seconds. The *CPUU* metric shows temporary use of all available CPU resources (some values close to 100), especially for *Dataset B*. However, the maximum resource consumption

in the cases studied was not a continuous state, but had few peaks (Figure 3).

Instantaneous maximum resource usage should not be a threat to the clustering process, but if the available resources are fully used for a longer period of time, the clustering could end in an error, e.g. due to a timeout. This risk is particularly high for memory-intensive algorithms (e.g. Agglomerative, Spectral), as in their case full RAM usage could cause the clustering to stop and generate an exception.

When comparing the algorithms analysed on the basis of the *CPUU* index, it is difficult to say which one is better. In the case of Dataset A, the maximum CPU consumption was slightly lower on average for the K-medians method, but the difference is not large and can hardly be considered significant. The results for Dataset B are very similar, but this is due to the fact that there were always times when the available CPU was fully utilised.

Some doubts may be raised when comparing the values of the *CD* and *CPUU* indicators. Looking directly at the *CPUU* values for K-medians, it could be argued that it is slightly better than the other approaches analysed. It should be noted that this is only a point maximum CPU load and does not take into account the clustering time. In fact, considering the *CD* metric, K-median clustering takes the longest time, so overall the CPU is used more than in the other approaches (due to the computational complexity). This means that the *CPUU* metric can be helpful in analysing the risk of CPU overload (and consequent interruption of clustering), but it does not determine the overall resource consumption.

In summary, the results obtained suggest that the *CPUU*

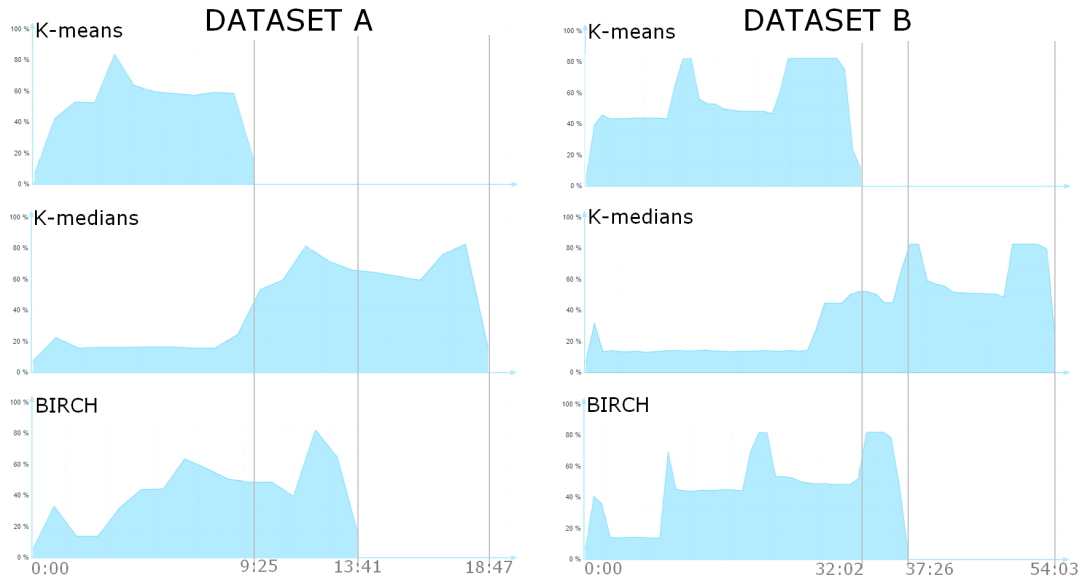


Figure 3. Comparison of CPU consumption for $k=7$.

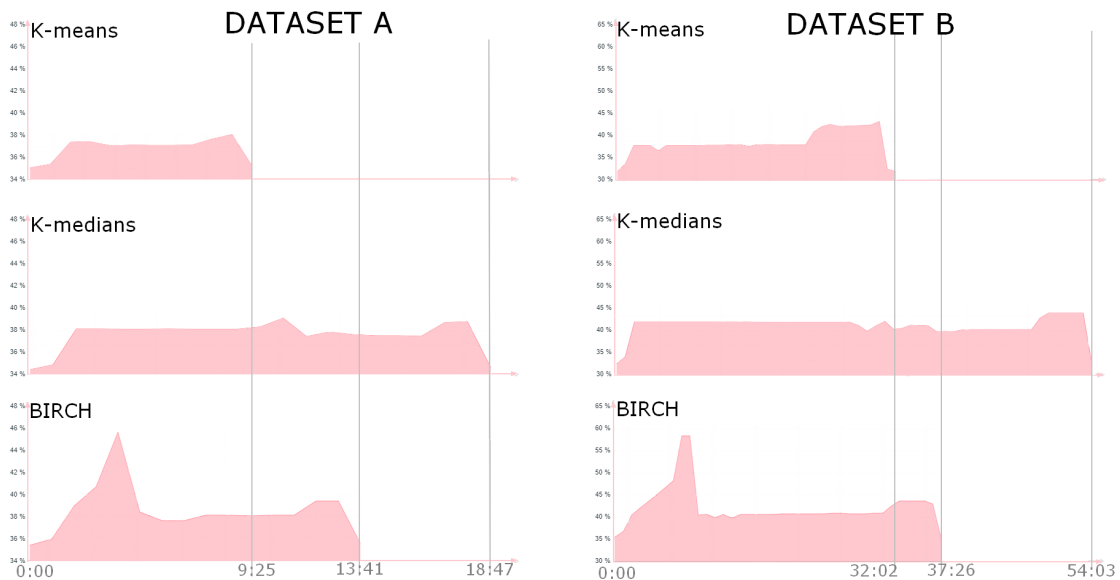


Figure 4. Comparison of memory (RAM) consumption for $k=7$.

metric needs to be modified. Its current design (the maximum value of resource usage) may not be sufficient to correctly quantify the level of CPU usage. By analysing the CPU load graphs, it is possible to see differences in CPU usage over time between the three algorithms, but these are not reflected in the *CPUU* values. Therefore, it would be necessary to change the design of the *CPUU* metric so that it covers the entire clustering process rather than just a single point value. In the simplest case, this could even be the mean and standard deviation of all the samples taken within the algorithm being analysed. This would make it possible to assess both the level of CPU usage and its variability over time.

In the case of the *MEMU*, the conclusions are similar, although not all of the operational memory was used (Figure 4). The values of this metric for both *K-...* algorithms tested are almost the same level within the dataset. The BIRCH method is slightly different. In its case, the values of the *MEMU*

metric are much higher, regardless of the dataset.

It can also be seen that the maximum memory consumption is higher for Dataset B, which is related to the higher number of customers to be grouped. Again, it is worth considering modifying this indicator to take account of changes in memory consumption over time. Admittedly, in the case of the algorithms analysed, this would probably not make much difference, but if approaches with more different memory consumption were compared, it could help in the interpretation of the results obtained.

Context-free quality

A summary of the calculated values for selected indicators of context-free clustering quality is presented in Table 6.

For these metrics, the number of clusters k used as an initialisation parameter has a very strong influence on the quality assessment. This means that comparisons between algorithms using these metrics should be made for the same k values. Furthermore, it is not appropriate to limit the anal-

Table 6. Context-free quality comparison

DATASET A									
k	SS			CHI			$IDBI$		
	K-means	K-medians	BIRCH	K-means	K-medians	BIRCH	K-means	K-medians	BIRCH
3	0.2585	0.2454	0.2534	20350.3024	15454.4377	19577.7296	0.3883	0.3544	0.3757
4	0.2050	0.2445	0.2327	17907.7865	11102.9185	14244.9426	0.4401	0.2733	0.3653
5	0.2118	0.2338	0.2357	15105.8067	8221.4264	11265.5824	0.3915	0.2885	0.2639
6	0.2301	0.2400	0.2236	13364.1800	7771.3484	9395.9682	0.4148	0.3281	0.2616
7	0.2305	0.1264	0.2252	12076.7233	5453.0872	8083.9662	0.3947	0.3216	0.2636
8	0.2347	0.0633	0.2244	10967.0855	5618.8102	7137.0875	0.3715	0.2650	0.2777
9	0.2407	0.0655	0.2267	10128.6545	4713.5173	6424.6934	0.3864	0.2679	0.2683
10	0.2399	0.1383	0.2293	9359.6237	5790.1597	5924.0808	0.3777	0.3149	0.2670
avr.	0.2314	0.1697	0.2314	13657.5203	8015.7132	10256.7563	0.3956	0.3017	0.2929
std.dev.	0.0169	0.0805	0.0098	3885.9425	3647.6465	4666.8281	0.0220	0.0328	0.0482
DATASET B									
3	0.1785	0.2417	0.2197	34602.8160	24398.9775	15654.9912	0.5158	0.2869	0.3389
4	0.1916	0.2407	0.2614	27486.3123	13081.7478	21159.7215	0.3975	0.3068	0.2835
5	0.2040	0.0937	0.2651	23750.6559	9050.2635	17240.0111	0.3776	0.3542	0.2820
6	0.2189	0.0944	0.2637	21143.8494	11811.8260	14605.5425	0.4088	0.2569	0.3117
7	0.2229	0.2385	0.2593	18996.5435	9216.9030	12490.3049	0.4007	0.3098	0.2960
8	0.2377	0.1066	0.2603	17753.8802	9210.8101	11104.9530	0.3929	0.2791	0.2915
9	0.2500	0.0656	0.2618	16547.3890	9789.0425	10053.4659	0.3976	0.3019	0.2946
10	0.2548	0.1409	0.2621	15597.0987	7369.8428	9263.3315	0.3947	0.3053	0.2953
avr.	0.2198	0.1528	0.2567	21984.8181	11741.1766	13946.5402	0.4107	0.3001	0.2992
std.dev.	0.0273	0.0753	0.0150	6442.1041	5411.0622	4026.9403	0.0434	0.0282	0.0185

ysis to a single number of clusters, as this may distort the conclusions. A good example would be the SS metric values. In this case, it is possible to identify the values of the parameter k for which each of the analysed clustering methods gives the best results. If the analysis had been done on just one value of k , it would not have been possible to see such differences.

When analysing the SS values, it is not possible to clearly identify the best clustering algorithm. The differences in the calculated values are so small and ambiguous that they cannot be the basis for reliable conclusions. This is an important finding as some studies [Sahinbas and Catak, 2022] have only used the Silhouette Score as a basis for evaluating different clustering techniques.

The situation is different for the CHI index. For all k analysed, the best values were obtained using the K-means algorithm, allowing us to conclude that the clusters generated in this way are denser and better separated. The BIRCH method ranks second for this indicator, and the worst results are obtained with K-median clustering.

Also, the $IDBI$ index indicated the best K-means clustering results, for both datasets. In this case, however, it is not possible to unambiguously indicate a further sequence. Depending on the value of the parameter k , better results were obtained with K-medians ($k = 5, 6, 7, 10$) or BIRCH clustering ($k = 3, 4, 8, 9$). These results again show that conclusions about the quality of context-free clustering depend on the choice of the parameter k .

Business context quality

From the point of view of the purposes of the study, in

terms of taking into account the business context of clustering when assessing the quality of the algorithms analysed, the $KMAX$, CSI and SDC metrics are crucial. These can be used to assess the practical usefulness of the resulting clusters. Ignoring business relevance aspects may result in a selected algorithm that is fast, computationally inexpensive and has good context-free quality values, but at the same time the resulting clusters are not applicable due to specific business requirements and needs.

The selected context-free quality indicators relate to the use of clustering to divide e-commerce customers into groups to be served a specific UI variant. This is an example of a practical personalisation of an online shop design that goes beyond the usual product recommendations. The contextual quality indicators analysed were selected based on identified business needs, including the number of clusters (which directly translates into the number of dedicated UI variants) and their size distribution. The values obtained are shown in Table 7. The $KMAX$ indicator is the number of clusters where the smallest cluster size exceeds the assumed threshold (thl). For the purposes of the study, the threshold was assumed to be 5% of the population of clustered e-commerce customers. With these settings, the K-means algorithm gave the best results for both datasets, but for Dataset B the same result was obtained using BIRCH clustering. Due to the specificity of dealing with dedicated UI variants, this indicator is of key importance. When comparing specific combinations [clustering method - number of clusters], compliance with the criterion described by the $KMAX$ indicator can be considered as a prerequisite for further analysis.

Table 7. Business context quality comparison

DATASET A									
k	$KMAX @ thl = 5\%$			CSI			SDC		
	K-means	K-medians	BIRCH	K-means	K-medians	BIRCH	K-means	K-medians	BIRCH
3	*	*	*	0.5452	0.8069	0.5581	16283.23	28028.69	16400.82
4	*		*	0.3123	0.9627	0.8612	4623.20	26458.76	20829.12
5	*		*	0.6631	0.9725	0.8612	8054.12	27507.00	19237.97
6	*			0.5623	0.9882	0.9025	5224.74	22936.08	18793.48
7	*			0.5598	0.9869	0.9444	4511.26	20894.35	18054.86
8	*			0.6278	0.9718	0.9737	5202.33	13919.89	17509.05
9	*			0.7531	0.9723	0.9737	5658.79	14827.52	16800.00
10				0.7620	0.9958	0.9737	5788.74	11468.95	16007.11
avr.	n/a	n/a	n/a	0.5982	0.9571	0.8811	6918.30	20755.15	17954.05
std.dev	n/a	n/a	n/a	0.1430	0.0617	0.1390	3941.14	6588.84	1619.62
DATASET B									
3	*	*	*	0.4434	0.8765	0.9064	20310.41	54705.05	75426.44
4	*		*	0.7143	0.9769	0.8787	23721.36	64010.39	46578.74
5	*		*	0.7274	0.9512	0.8787	22140.09	51964.94	45095.64
6				0.9326	0.9198	0.9689	24584.99	40905.23	43632.11
7				0.9333	0.9804	0.9689	23562.84	51338.16	41806.17
8				0.9190	0.9582	0.9689	19093.42	41633.04	39468.74
9				0.9081	0.9900	0.9689	14987.02	25739.58	37577.63
10				0.9083	0.9889	0.9689	14279.97	40464.48	35864.73
avr.	n/a	n/a	n/a	0.8108	0.9552	0.9388	20335.01	46345.11	45681.28
std.dev	n/a	n/a	n/a	0.1740	0.0395	0.0430	3964.37	11646.15	12568.49

The CSI indicator shows the spread of cluster sizes. From a business point of view, we would expect the spread to be as small as possible, so that a similar number of customers are served with dedicated user interface variants. Of course, perfectly equal sizes cannot be expected, but those algorithms that produce clusters that are as similar in size as possible can be considered preferable. For the datasets analysed, the best CSI values were obtained for the K-means algorithm compared to both K-medians and BIRCH. Furthermore, the analysis of this metric makes it possible to determine the optimal number of clusters when the decision is based on minimising its value. Noteworthy in this context are the results of the K-means clustering for $k=4$ and Dataset A, which gave by far the lowest CSI value.

The SDC metric measures the variation in cluster size relative to the average and relates to the same business requirement as the CSI metric. Preference should be given to algorithms for which SDC has smaller values. In the case of the study described above, K-means undoubtedly performed best. If this metric were to be used as a criterion for selecting specific clustering parameters for serving dedicated UI variants, the clustering at $k=7$ and Dataset A, which gives the lowest SDC value of all the options analysed, deserves attention.

4.3 Selection of the clustering method

The results of the experimental study allowed a practical validation of the proposed cluster quality framework, with a particular focus on indicators derived from the business context of the use of the results. Two similar algorithms,

K-means and K-medians, were selected for comparison. The study also included the BIRCH method, which has a different approach to clustering. An additional challenge is to aggregate the sub-scores (related to individual quality measures) to identify the solution that best fits the business specificity of serving a multi-variant UI in e-commerce. Two approaches to solving this problem are presented later in this subsection.

The simplest comparison of clustering methods can use ranks. Based on the results, the techniques studied can be ranked, with a value of **1** assigned to the best algorithm, **2** to the middle algorithm, and **3** to the worst algorithm. Several algorithms could be given the same rank if the obtained values of the applied metric were similar. The rank values can be used for an overall multi-criteria comparison of the analyzed algorithms. In addition, the weights assigned to the quality indicators can be used to determine the relevance of the metric to a specific business application (nevertheless, to simplify the example given, the same weights were assumed for all the deciding criteria).

Nine indicators were used in the analysis and the assigned ranks for each clustering method are shown in Table 8. The subjective weighting of each indicator is intended to provide an approach to aggregating the sub-scores. For the assumptions used in the study, the K-means algorithm proved to be significantly better, receiving a rank of **1** in most cases. Only for the CPUU indicator did the K-medians method perform better.

Another way to solve the problem of aggregating sub-measures is to use one of the Multiple Criteria Decision

Table 8. Aggregate comparison of clustering methods

Metric	Weight (w_j)	K-means	K-medians	BIRCH
CD	1	1	3	2
CPUU	1	2	1	3
MEMU	1	1	1	3
SS	1	1	3	1
CHI	1	1	3	2
IDB	1	1	2	2
KMAX	1	1	3	2
CSI	1	1	3	2
SDC	1	1	3	2
Total weighted		10	22.0	19.0

Analysis (MCDA) methods. TOPSIS was selected as an exemplary approach from this group. It's a technique that aims to identify the preferred choice by comparing alternatives against ideal and anti-ideal solutions. In the analyzed case, all combinations [clustering method - number of clusters] that satisfy the critical requirements (also derived from business specifics) can be considered as alternatives. The *KMAX* metric can be such a critical condition, as it refers to the minimum reasonable number of users in a cluster. With this assumption, 11 alternatives can be found for Dataset A and 7 for Dataset B (Table 7). In turn, the other measures identified in the framework described can be treated as decision criteria with specific weights.

The presentation of the application of the adopted TOPSIS method for selecting a clustering method for serving dedicated UI variants was based on Dataset A.

The starting point is the decision matrix $D = \{d_{ij}\}$, where alternatives A_i , $i = 1, 2, \dots, 11$ are presented in rows and the criteria X_j , $j = 1, 2, \dots, 8$ are presented in columns. The corresponding values of the metrics shown in Tables 5- 7 were taken as d_{ij} . It is worth noting that it was necessary to make the interpretation of the values of the various metrics more consistent. For all metrics, it was assumed that the lower the value of the metric means the higher the rating of the analyzed alternative. Accordingly, the values of the metrics for which the results were to be maximized (*SS*, *CHI*, *IDBI*) were inverted.

The next step is to normalize the decision matrix. The normalization with the Euclidean norm was used as a basis for this operation:

$$r_{ij} = \frac{d_{ij}}{\sqrt{\sum_i d_{ij}^2}} \quad (2)$$

giving the normalized decision matrix $R = \{r_{ij}\}$.

The weighted decision matrix $V = \{v_{ij}\}$ was then calculated using the same weights for each decision factor ($\forall w_j = 1$):

$$v_{ij} = w_j * r_{ij} \quad (3)$$

After preliminary data preparation, the ideal solution A^+ and the negative ideal solution A^- were calculated for each decision factor j :

$$A_j^+ = \min_i(v_{ij}), A_j^- = \max_i(v_{ij}) \quad (4)$$

Then the separation measures S_i^+ and S_i^- were calculated for each alternative i :

$$S_i^+ = \sqrt{\sum_j (v_{ij} - A_j^+)^2}, S_i^- = \sqrt{\sum_j (v_{ij} - A_j^-)^2} \quad (5)$$

Finally, the relative closeness C_i to the ideal solution was calculated for each alternative:

$$C_i = \frac{S_i^-}{(S_i^- + S_i^+)} \quad (6)$$

The alternatives, ranked in descending order of C_i value, allow an aggregated assessment of the available clustering options, taking into account all the criteria adopted for the analysis (Table 9).

Table 9. Evaluation of methods using the TOPSIS approach

Rank	Method(k)	S_i^+	S_i^-	C_i
1	K-means(4)	0.0911	0.6387	0.8752
2	K-means(6)	0.1629	0.5778	0.7801
3	K-means(7)	0.1818	0.5882	0.7638
4	K-means(5)	0.2011	0.5215	0.7217
5	K-means(8)	0.2343	0.5543	0.7029
6	K-means(9)	0.2918	0.5417	0.6499
7	K-means(3)	0.2723	0.4466	0.6212
8	BIRCH(3)	0.2950	0.3975	0.5740
9	BIRCH(4)	0.4490	0.2670	0.3729
10	BIRCH(5)	0.4711	0.2480	0.3449
11	K-medians(3)	0.5973	0.1928	0.2440

The advantage of the K-means method in both aggregation approaches can be explained by several factors. The main one is the difference in computational complexity, especially for large and multidimensional datasets. For K-means and K-medians, differences in the operations performed in the clustering process are important. Sorting, as required in K-medians, generally has a higher computational cost than summing, as used in K-means. Admittedly, this fact was not reflected in the values of the CPUU indicator in the tests carried out (especially for Dataset A), but this is due to the specific nature of this measure, which is point-based and additionally sampled every 60 seconds. The analysis of the CD index values clearly shows the higher computational complexity of the K-median method, as the total clustering time is significantly longer. The study confirmed that the sorting process, which involves comparing items and reordering them, can be inefficient for large data sets or multidimensional data. BIRCH, on the other hand, requires the construction of a CF tree (Clustering Feature Tree) data structure, which involves scanning the dataset to construct an initial clustering structure. This initialisation step can be computationally expensive, especially for large datasets. As a result, the BIRCH took longer to cluster than the K-means method, although it was shorter than the K-medians. It should be noted that BIRCH requires more memory than K-means or K-medians because it retains the CF-tree data structure.

Another reason why K-means was found to be the best method may be the specificity of the data sets. K-means assumes that clusters are spherical and approximately equal in

size, and it works best with data that is approximately normally distributed. The results obtained indicate that this is the nature of the datasets containing information on the customer behaviour of the online shops analysed. These features influence the results of context-free and contextual quality of clustering. If the clusters had different shapes, or if the data were non-normal with many outliers, the K-medians method might be expected to give better results.

Another factor influencing the results obtained is the number of clusters analysed. Due to the requirements of the business context, k took a value between 3 and 10. Had it been necessary to generate a significantly higher number of clusters, the results of the method comparison (especially in terms of resource usage and context-free quality) might have shown a better efficiency of the BIRCH method.

The research also indicated that some of the metrics should be modified. This mainly concerns the *CPUU* and *MEMU* metrics, which in their current form result in the loss of some information about the characteristics of the clustering algorithms being studied. As one of the risks associated with these metrics is the inability to complete clustering, e.g. due to insufficient memory to store the processed data, a potential direction for improving these metrics could be to try to determine the maximum dataset that can be correctly processed with given system resources. In addition, it is worth considering the specificity of cloud servers, where resource consumption has a certain financial cost. In this case, instead of analysing CPU and memory usage, the cost of clustering expressed in currency units can be used as a metric. In such a case, additional attention should be paid to finding the optimal configuration of cloud servers so that the resources match the specifics of the clustering method and to avoid paying for ordered but unused computing capacity. Moreover, the results presented in the paper on resource consumption during clustering could be helpful in selecting cloud server parameters.

4.4 Analysis of the impact of the proposed framework

TOPSIS was used to verify the impact of applying the proposed framework on the evaluation of the different clustering approaches. In a first step, the ranking of the clustering options was calculated (following the principles described in Section 4.3), using only the context-free quality metrics (*SS*, *CHI*, *IDBI*) as decision factors. Such an operation is similar to the most commonly used criteria for evaluating clustering algorithms when performance issues and business context requirements are not taken into account. The results obtained can therefore be considered as a benchmark and can be interpreted as a ranking without applying the proposed quality framework. The evaluation of the clustering alternatives using this approach is presented in Table 10. Analysis of the results shows that the best option would be [K-means; $k=3$]. This was followed by the [BIRCH; $k=3$], [K-means; $k=4$] and [K-medians; $k=3$] approaches. It can therefore be concluded that the high rankings went to alternatives with a small number of outcome clusters, and that each of the methods analysed had a representative in the top positions.

In the next step of the study, the set of context-free metrics

Table 10. TOPSIS evaluation based on context-free metrics

Rank	Method(k)	S_i^+	S_i^-	C_i
1	K-means(3)	0.0336	0.2532	0.8829
2	BIRCH(3)	0.0442	0.2403	0.8446
3	K-means(4)	0.0762	0.2427	0.7610
4	K-medians(3)	0.0892	0.1844	0.6740
5	K-means(5)	0.0972	0.1910	0.6629
6	K-means(6)	0.1110	0.1852	0.6253
7	BIRCH(4)	0.1047	0.1698	0.6186
8	K-means(7)	0.1431	0.1597	0.5273
9	K-means(8)	0.1789	0.1355	0.4309
10	K-means(9)	0.2065	0.1422	0.4079
11	BIRCH(5)	0.2348	0.0605	0.2049

was extended to include resource consumption metrics - *CD*, *CPUU*, *MEMU*. The results are shown in Table 11. Intu-

Table 11. TOPSIS evaluation based on context-free and resource usage metrics

Rank	Method(k)	S_i^+	S_i^-	C_i
1	K-means(3)	0.0565	0.3455	0.8594
2	K-means(4)	0.0910	0.3312	0.7844
3	K-means(5)	0.0987	0.3028	0.7541
4	K-means(6)	0.1161	0.3016	0.7220
5	BIRCH(3)	0.1165	0.2842	0.7092
6	K-means(7)	0.1431	0.2976	0.6752
7	K-means(8)	0.1851	0.2668	0.5905
8	BIRCH(4)	0.1576	0.2212	0.5840
9	K-means(9)	0.2115	0.2740	0.5644
10	K-medians(3)	0.2605	0.1912	0.4233
11	BIRCH(5)	0.2579	0.1678	0.3942

itively, it might have been expected that the additional metrics would positively affect the positions of the alternatives applying K-means clustering, while negatively affecting the alternative applying K-medians clustering. The results obtained confirmed this hypothesis. While the best alternative [K-means; $k=3$] has not changed, there have been significant changes in the next places. In particular, the [BIRCH; $k=3$] and [K-medians; $k=3$] combinations perform significantly worse, which is a clear result of the superiority of the K-means method in terms of computational complexity and the server resources required.

Interesting conclusions can be drawn by analyzing the case of using all three characteristics included in the proposed framework (Table 9) and comparing it with the application of one (Table 10) and two (Table 11) characteristics. The addition of context-aware clustering quality metrics makes the combination [K-means; $k=4$] the best option, while the previously best-rated alternative ([K-means; $k=3$]) drops to 7th place. This means that consideration of the business context can be important when choosing the clustering method and parameters for results applications to serve dedicated UI variants. It is noteworthy that the option found to be the best using all the metrics from the quality framework described above exactly matches the expert-based decisions made in the previous studies ([Wasilewski and Kolaczek, 2024]), where the effectiveness of e-commerce multi-variant UIs was analyzed based on 4 clusters obtained after applying the K-

means method.

5 Conclusions and future work

Personalisation in e-commerce is a trend that is shaping the development of today's online shops. One way to ensure a better UX is to tailor the design to the requirements, needs and behaviour of customers. This is currently done to a limited extent, mainly in the form of product recommendations. However, a more comprehensive approach could be considered, where a multi-variant interface is offered, moving away from the *one size fits all* approach. However, the design of multiple interface options requires an appropriate grouping of customers. This can be done using traditional segmentation methods, but also using clustering techniques, which are becoming increasingly common. However, in order to apply any of the clustering algorithms, it is necessary to make a rational assessment of the available options, taking into account many aspects, including the business context.

To address the problem of multidimensional evaluation of clustering quality, the quality framework has been proposed that considers three main characteristics: resource usage, context-free quality and business context-sensitive quality. The first two are usually analysed during clustering algorithm selection, but the third is sometimes neglected. Therefore, three metrics are proposed to test the degree of fit between clustering results and business requirements and constraints when serving dedicated UI variants in e-commerce.

An experimental study was then carried out with two different learning datasets and the three clustering algorithms, K-means, K-medians, and BIRCH, were compared. This allowed verification of the proposed approach to selecting the clustering algorithm and an indication of which method should be chosen to perform the clustering, given the specified input data and available options, if the results were to be used to serve dedicated user interface variants to e-commerce customers.

The practical verification of the proposed quality framework has also identified weaknesses. The development and re-development of such areas will allow the proposed approach to be refined in the future. In addition, further work should be aimed at making the model even more flexible by opening it up to additional characteristics and metrics to allow even better evaluation and comparison of clustering algorithms for specific business purposes.

Declarations

Acknowledgements

Clustering was performed using the *AIM²* platform developed by Fast White Cat S.A., Poland.

Author's Contributions

The research described in this paper was conducted by Adam Wasilewski.

Competing Interests

The author declares no conflict of interest.

Availability of data and materials

The data that was used in the study has been made available in a public repository: <https://doi.org/10.7910/DVN/5YUE8N>

References

- Aksoy, N. C., Kabadayi, E. T., Yilmaz, C., and Alan, A. K. (2023). Personalization in marketing: How do people perceive personalization practices in the business world? *Journal of Electronic Commerce Research*, 24(4):269–297. Available at: <http://www.jecr.org/node/693>.
- Al-Kilidar, H., Cox, K., and Kitchenham, B. (2005). The use and usefulness of the iso/iec 9126 quality standard. In *2005 International Symposium on Empirical Software Engineering, 2005.*, pages 7–pp. IEEE. DOI: 10.1109/IS-ESE.2005.1541821.
- Albert, B., Tullis, T., and Tadesco, D. (2010). *Beyond the Usability Lab*. Elsevier. DOI: 10.1016/C2009-0-19827-6.
- Amini, A. and Haughton, M. (2023). A mathematical optimization model for cluster-based single-depot location-routing e-commerce logistics problems. *Supply Chain Analytics*, 3. DOI: 10.1016/j.sca.2023.100019.
- Amna Altaf, Adnen El Amraoui, F. D. and Lecoutre, C. (2023). Applications of artificial intelligence in cross docking: A systematic literature review. *Journal of Computer Information Systems*, 63(5):1280–1300. DOI: 10.1080/08874417.2022.2143455.
- Calinski, T. and Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics - Theory and Methods*, 3:1–27. DOI: 10.1080/03610927408827101.
- Camilleri, M. A. (2017). *Market Segmentation, Targeting and Positioning*. Springer. DOI: 10.1007/978-3-319-49849-2.
- Chen, R., Jia, S., and Meng, Q. (2023). Dynamic container drayage booking and routing decision support approach for e-commerce platforms. *Transportation Research Part E: Logistics and Transportation Review*, 177. DOI: 10.1016/j.tre.2023.103220.
- Cui, H., Niu, S., Li, K., Shi, C., Shao, S., and Gao, Z. (2021). A k-means++ based user classification method for social e-commerce. *Intelligent Automation & Soft Computing*, 28:277–291. DOI: 10.32604/iasc.2021.016408.
- Dasgupta, S., Frost, N., Moshkovitz, M., and Rashtchian, C. (2020). Explainable k-means and k-medians clustering. In *Proceedings of the 37th International Confer-*

- ence on Machine Learning, ICML'20. JMLR.org. DOI: 10.48550/arXiv.2002.12538.
- Davies, D. L. and Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2):224–227. DOI: 10.1109/TPAMI.1979.4766909.
- Desaid, D. (2019). 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. *Advances in Intelligent Systems and Computing*, pages 3–19. DOI: 10.1007/978-981-13-1274-8₁.
- Dolnicar, S., Grün, B., and Leisch, F. (2018). *Market Segmentation Analysis*. Springer Singapore. DOI: 10.1007/978-981-10-8818-6.
- Dunn, J. C. (1973). A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 3(3):32–57. DOI: 10.1080/01969727308546046.
- Estdale, J. and Georgiadou, E. (2018). Applying the iso/iec 25010 quality models to software product. In *Systems, Software and Services Process Improvement: 25th European Conference, EuroSPI 2018, Bilbao, Spain, September 5-7, 2018, Proceedings 25*, pages 492–503. Springer. DOI: 10.1007/978-3-319-97925-0₄₂.
- Faraone, M., Gorgoglione, M., Palmisano, C., and Panniello, U. (2012). Using context to improve the effectiveness of segmentation and targeting in e-commerce. *Expert Systems with Applications*, 39(9):8439–8451. DOI: 10.1016/j.eswa.2012.01.174.
- Fontanini, A. D. and Abreu, J. (2018). A data-driven birch clustering method for extracting typical load profiles for big data. In *2018 IEEE Power & energy society general meeting (PESGM)*, pages 1–5. IEEE. DOI: 10.1109/PESGM.2018.8586542.
- Gomes, M. and Meisen, T. (2023). A review on customer segmentation methods for personalized customer targeting in e-commerce use cases. *Information Systems and e-Business Management*, 21:1–44. DOI: 10.1007/s10257-023-00640-4.
- Guo, G. and Altrjman, C. (2022). E-commerce customer segmentation method under improved k-means algorithm. In Sugumaran, V., Sreedevi, A. G., and Xu, Z., editors, *Application of Intelligent Systems in Multi-modal Information Analytics*, pages 1083–1089. Springer International Publishing. DOI: 10.1007/978-3-031-05484-6₁₄₈.
- Han, J., Kamber, M., and Pei, J. (2011). *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 3rd edition. DOI: 10.1016/C2009-0-61819-5.
- Han, L., Fang, J., Zheng, Q., George, B. T., Liao, M., and Hossin, M. A. (2024). Unveiling the effects of livestream studio environment design on sales performance: A machine learning exploration. *Industrial Marketing Management*, 117:161–172. DOI: 10.1016/j.indmarman.2023.12.021.
- Hicham, N. and Karim, S. (2022). Analysis of unsupervised machine learning techniques for an efficient customer segmentation using clustering ensemble and spectral clustering. *International Journal of Advanced Computer Science and Applications*, 13(10). DOI: 10.14569/IJACSA.2022.0131016.
- Hjort, K., Lantz, B., Ericsson, D., and Gattorna, J. (2016). *Customer Segmentation Based on Buying and Returning Behaviour: Supporting Differentiated Service Delivery in Fashion E-Commerce*, pages 153–169. Palgrave Macmillan UK, London. DOI: 10.1057/9781137541253₁₄.
- Hwang, C.-L. and Yoon, K. (1981). *Multiple attribute decision making: methods and applications a state-of-the-art survey*. Springer Science & Business Media. DOI: 10.1007/978-3-642-48318-9.
- John, J., Shobayo, O., and Ogunleye, B. (2023). An exploration of clustering algorithms for customer segmentation in the uk retail market. *Analytics*, 2:809–823. DOI: 10.3390/analytics2040042.
- Koehn, D., Lessmann, S., and Schaal, M. (2020). Predicting online shopping behaviour from clickstream data using deep learning. *Expert Systems with Applications*, 150. DOI: 10.1016/j.eswa.2020.113342.
- Kopel, M., Sobacki, J., and Wasilewski, A. (2013). Automatic web-based user interface delivery for soa-based systems. *Computational Collective Intelligence*, 8083:110–119. DOI: 10.1007/978-3-642-40495-5₁₂.
- Li, P., Wang, C., Wu, J., and Madlenak, R. (2022). An e-commerce customer segmentation method based on rfm weighted k-means. In *Proceedings - 2022 International Conference on Management Engineering, Software Engineering and Service Sciences, ICMSS 2022*, page 61 – 68. DOI: 10.1109/ICMSS55574.2022.00017.
- Lorbeer, B., Kosareva, A., Deva, B., Softić, D., Ruppel, P., and Küpper, A. (2017). Variations on the clustering algorithm birch. *Big Data Research*, 11. DOI: 10.1016/j.bdr.2017.09.002.
- Ma, J. (2022). E-commerce customer segmentation based on rfm model. In Hung, J. C., Yen, N. Y., and Chang, J.-W., editors, *Frontier Computing*, pages 926–931, Singapore. Springer Nature Singapore. DOI: 10.1007/978-981-16-8052-6₁₁₈.
- Mashalah, H. A., Hassini, E., Gunasekaran, A., and Bhatt (Mishra), D. (2022). The impact of digital transformation on supply chains through e-commerce: Literature review and a conceptual framework. *Transportation Research Part E: Logistics and Transportation Review*, 165. DOI: 10.1016/j.tre.2022.102837.
- Maulana, A. D., Ningsih, A. K., and Abdillah, G. (2023). Consumer segmentation using k-medians algorithm on transaction data based on lrmp (length, recency, frequency, monetary, periodicity). *Enrichment: Journal of Multidisciplinary Research and Development*, 1(8):477–483. DOI: 10.55324/enrichment.v1i8.70.
- Meena, P., Kumar, C., and Puri, S. (2023). Customer segmentation and behavioral systems through influential effective elements: An e-satisfaction analysis using machine learning. In *AIP Conference Proceedings*, volume 2782. DOI: 10.1063/5.0154287.
- Nanayakkara, P. R., Jayalath, M. M., Thibbotuwawa, A., and Perera, H. N. (2022). A circular reverse logistics framework for handling e-commerce returns. *Cleaner Logistics*

- and Supply Chain, 5. DOI: 10.1016/j.clscn.2022.100080.
- Nawara, D. and Kashef, R. (2021). Deploying different clustering techniques on a collaborative-based movie recommender. In *2021 IEEE International Systems Conference (SysCon)*, pages 1–6. DOI: 10.1109/SysCon48628.2021.9447139.
- Nguyen, T. T., Phan, T. C., Pham, H. T., Nguyen, T. T., Jo, J., and Nguyen, Q. V. H. (2023). Example-based explanations for streaming fraud detection on graphs. *Information Sciences*, 621:319–340. DOI: 10.1016/j.ins.2022.11.119.
- Nurma Sari, J., Nugroho, L., Ferdiana, R., and Santosa, P. (2016). Review on customer segmentation technique on e-commerce. *Advanced Science Letters*, 22:3018–3022. DOI: 10.1166/asl.2016.7985.
- Okon, E., Eke, B., and Asagba, P. (2018). An improved online book recommender system using collaborative filtering algorithm. *International Journal of Computer Applications*, 179. DOI: 10.13140/RG.2.2.24240.46086.
- Ooi, K.-B. O., Tan, G. W.-H., Mostafa Al-Emran, M., and Al-Sharafi, M. A. a. (2023). The potential of generative artificial intelligence across disciplines: Perspectives and future directions. *Journal of Computer Information Systems*, 0(0):1–32. DOI: 10.1080/08874417.2023.2261010.
- Papamichail, G. P. and Papamichail, D. P. (2007). The k-means range algorithm for personalized data clustering in e-commerce. *European Journal of Operational Research*, 177(3):1400–1408. DOI: 10.1016/j.ejor.2005.04.011.
- Punhani, R., Arora, V., Sabitha, A. S., and Shukla, V. K. (2020). Segmenting e-commerce customer through data mining techniques. *Journal of Physics: Conference Series*, 1714:1–12. DOI: 10.1088/1742-6596/1714/1/012026.
- Punhani, R., Arora, V., Sabitha, S., and Shukla, V. K. (2021). Application of clustering algorithm for effective customer segmentation in e-commerce. In *Proceedings of the 2021 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*, pages 149–154. IEEE. DOI: 10.1109/ICCIKE51210.2021.9410713.
- Rajput, L. and Singh, S. N. (2023). Customer segmentation of e-commerce data using k-means clustering algorithm. In *2023 13th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, pages 658–664. DOI: 10.1109/Confluence56041.2023.10048834.
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65. DOI: 10.1016/0377-0427(87)90125-7.
- Sahinbas, K. and Catak, F. O. (2022). Customer segmentation in the retail sector: A data analytics approach. In *2022 14th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, pages 174–178. DOI: 10.1109/IHMSC55436.2022.00048.
- Shen, X. (2023). E-commerce user recommendation algorithm based on social relationship characteristics and improved k-means algorithm. *International Journal of Computational Intelligence Systems*, 16. DOI: 10.1007/s44196-023-00321-7.
- Sihombing, P. (2021). Implementation of k-means and k-medians clustering in several countries based on global innovation index (gii) 2018. *Advance Sustainable Science, Engineering and Technology*, 3:0210107. DOI: 10.26877/asset.v3i1.8461.
- Solichin, A. and Wibowo, G. (2022). Customer segmentation based on recency frequency monetary (rfm) and user event tracking (uet) using k-means algorithm. In *Proceeding - IEEE 8th Information Technology International Seminar, ITIS 2022*, page 257 – 262. DOI: 10.1109/ITIS57155.2022.10009981.
- Song, Y. W. G., Lim, H. S., and Oh, J. (2021). “we think you may like this”: An investigation of electronic commerce personalization for privacy-conscious consumers. *Psychology & Marketing*, 38(10):1723–1740. DOI: 10.1002/mar.21501.
- Su, Q. and Chen, L. (2015). A method for discovering clusters of e-commerce interest patterns using click-stream data. *Electronic Commerce Research and Applications*, 14(1):1–13. DOI: 10.1016/j.elerap.2014.10.002.
- Tabianan, K., Velu, S., and Ravi, V. (2022). K-means clustering approach for intelligent customer segmentation using customer purchase behavior data. *Sustainability*, 14(12). DOI: 10.3390/su14127243.
- Tsao, Y.-C., Chen, Y.-K., Chiu, S.-H., Lu, J.-C., and Vu, T.-L. (2022). An innovative demand forecasting approach for the server industry. *Technovation*, 110:102371. DOI: 10.1016/j.technovation.2021.102371.
- Tsao, Y.-C., Liu, Y.-H., Vü, L., and Fang, I.-W. (2023). Intelligent design suggestion and sales forecasting for new products in the apparel industry. *Fibres & Textiles in Eastern Europe*, 31:30–38. DOI: 10.2478/ftce-2023-0052.
- Wang, G., Zhang, X., Tang, S., Wilson, C., Zheng, H., and Zhao, B. (2017). Clickstream user behavior models. *ACM Transactions on the Web*, 11:1–37. DOI: 10.1145/3068332.
- Wasilewski, A. (2019). Integration challenges for outsourcing of logistics processes in e-commerce. In *Asian Conference on Intelligent Information and Database Systems*. DOI: 10.1007/978-3-030-14132-5_29.
- Wasilewski, A. (2024). Functional framework for multivariate e-commerce user interfaces. *Journal of Theoretical and Applied Electronic Commerce Research*, 19(1):412–430. DOI: 10.3390/jtaer19010022.
- Wasilewski, A. and Kolaczek, G. (2024). One size does not fit all: Multivariate user interface personalization in e-commerce. *IEEE Access*, 12(2024):65570–65582. DOI: 10.1109/ACCESS.2024.3398192.
- Wasilewski, A. and Przyborowski, M. (2023). Clustering methods for adaptive e-commerce user interfaces. In *International Joint Conference on Rough Sets*, pages 511–525. Springer. DOI: 10.1007/978-3-031-50959-9_35.
- Wu, R.-S. and Chou, P.-H. (2011). Customer segmentation of multiple category data in e-commerce using a soft-clustering approach. *Electronic Commerce Research and Applications*, 10(3):331–341. DOI: 10.1016/j.elerap.2010.11.002.
- Wu, T. and Liu, X. (2020). A dynamic interval type-2 fuzzy customer segmentation model and its application in e-commerce. *Applied Soft Computing*, 94:106366. DOI: 10.1016/j.asoc.2020.106366.

- Xiao, B. and Benbasat, I. (2007). E-commerce product recommendation agents: Use, characteristics, and impact. *MIS Q.*, 31:137–209. DOI: 10.2307/25148784.
- Zare, H. and Emadi, S. (2020). Determination of customer satisfaction using improved k-means algorithm. *Soft Computing*, 24(22):16947 – 16965. DOI: 10.1007/s00500-020-04988-4.
- Zhang, J., Wu, J., and Gao, C. (2022). Consumption behavior analysis of e-commerce users based on k-means algorithm. *Journal of Network Intelligence*, 7(4):935 – 942. Available at: <https://bit.kuas.edu.tw/~jni/2022/vol17/s4/09.JNI0380.pdf>.
- Zhao, H.-H., Luo, X.-C., Ma, R., and Lu, X. (2021). An extended regularized k-means clustering approach for high-dimensional customer segmentation with correlated variables. *IEEE Access*, 9:48405–48412. DOI: 10.1109/ACCESS.2021.3067499.
- Zheng, K., Huo, X., Jasimuddin, S., Zhang, J. Z., and Battaia, O. (2023). Logistics distribution optimization: Fuzzy clustering analysis of e-commerce customers' demands. *Computers in Industry*, 151. DOI: 10.1016/j.compind.2023.103960.