Spoor, Jan Michael

Improving customer segmentation via classification of key accounts as outliers

Journal of Marketing Analytics

Provided in Cooperation with:
Springer Nature

Suggested Citation: Spoor, Jan Michael (2022) : Improving customer segmentation via classification of key accounts as outliers, Journal of Marketing Analytics, ISSN 2050-3326, Palgrave Macmillan, London, Vol. 11, Iss. 4, pp. 747-760, https://doi.org/10.1057/s41270-022-00185-4

Journal of Marketing Analytics (2023) 11:747–760
https://doi.org/10.1057/s41270-022-00185-4

ORIGINAL ARTICLE

Improving customer segmentation via classification of key accounts as outliers

Jan Michael Spoor[1]

Revised: 5 June 2022 / Accepted: 12 September 2022 / Published online: 30 September 2022

Abstract

Customer segmentation and key account management are important use cases for clustering algorithms. Here, a data set of a Portuguese wholesaler for food and household supplies is used as an exemplary application. To increase the quality of the analysis, a two-stage approach is proposed. First, key accounts are filtered by a density-based outlier detection. Second, a Gaussian Mixture Model (GMM) is applied to cluster smaller customers. This two-stage approach is aligned with the business implications of key accounts as outstanding and very differently behaving customers as well as with the core idea of an ABC analysis. Also, the exclusion of key accounts corresponds to the definition of outliers as the results of a different underlying mechanism. Using this two-stage approach shows better clustering results compared to using a one-stage approach applying only a GMM. Therefore, it is concluded that density-based detection of key accounts followed by a clustering using a GMM is beneficial for customer segmentation within B2B applications.

Keywords Clustering algorithms · Anomaly detection · Customer segmentation · Marketing management

Introduction

The aim of a cluster analysis within an explorative data analysis is the set-up of groups out of a data set (James et al. 2013). Cluster analyses are used in a multitude of applications in science, medicine, and business management tasks (Murphy 2012) with a notable application in customer segmentation (Jensen 2001). Cluster analysis is often used as a tool in setting up customer segments. Typically used algorithms are K-Means or Gaussian Mixture Models (GMMs) solved by an Expectation-Maximization Algorithm (EM Algorithm) (Jensen 2001).

Due to the increasing importance of e-commerce, customer segmentations are used to group users in clusters based on shopping and usage behavior in online stores in order to target individual advertising to each group (Berkhin 2006). Also, in more traditional businesses, customer segmentation is an important tool for success and a core competence of the marketing departments (Homburg and Krohmer 2006). An important task of the customer segment management is the management of key accounts in Business-To-Business (B2B) relationships. Key accounts are a group of very important customers which distinguish themselves from regular customers by measures like revenue and the involved risk if the business relation is weakened (Sidow 2000).

In this contribution, the proposed idea is to include domain-specific knowledge in the data analysis by viewing key accounts as anomalies. Anomaly detection, often interchangeably called outlier detection (Aggarwal 2013) and closely related to novelty detection (Chandola et al. 2009), is the task of finding data points not compatible with the assumed normal model of a data set (Aggarwal 2013). The anomalous data points deviate substantially from the normal model of the data (Mehrotra et al. 2017). Outliers are well known to distort data analysis and results, which has led to in the development of robust statistical methods less sensitive to the influence of outliers (Huber and Ronchetti 2009). Due to their deviation from regular customers regarding revenue and complexity, key accounts fit the description of data points deviating from the norm. Therefore, this paper defines an approach to conducting cluster analysis in order to identify the key accounts using methods of anomaly detection. The proposed approach utilizes a two-step clustering where the first step is the identification of key accounts by

✉ Jan Michael Spoor
jan.spoor@kit.edu

[1] Institut für Informationsmanagement im Ingenieurwesen (IMI), Karlsruhe Institute of Technology, Kriegsstraße 77, 76133 Karlsruhe, Baden-Württemberg, Germany

low-density areas in the data. Key accounts are then removed from the data set and the customer segmentation is conducted using a GMM. Simultaneously, the prior exclusion of key accounts will result in less distorted data, a better explanatory value of the key account segment, and more distinct cluster amount recommendations, while the GMM provides a fuzzy assignment instead of a strict segmentation. Therefore, this method addresses common limitations of customer segmentation using methods of cluster analysis.

The state of the art in customer segmentation is briefly presented in the section "State-of-the-art approaches and methods in customer segmentation". In the "Application of key accounts in customer segmentation and cluster analysis" section, the definitions of key accounts are discussed, as well as the problems in creating a distinct key accounts assignment using common measures. Subsequently, in the "Methodology" section, the methodology is explained and introduced, in particular the applied approach separating key accounts as anomalies within the data, the used clustering algorithms, and the information criteria for model selection. Conclusively in the section "Use-case: customer segmentation for a B2B wholesaler of food & household supplies", the method is applied to a real data set, and the implications for the cluster quality of the new approach are discussed by comparing the results to a previously conducted study.

State-of-the-art approaches and methods in customer segmentation

Commonly used methods for customer segmentation in marketing are distance-based algorithms, i.e., K-Means, and hierarchical algorithms, often agglomerative approaches, i.e., Ward. However, notable limitations of distance-based methods within practical application are the inability to clearly determine a cluster amount (Brudvig et al. 2019) and the assumption of isotropic cluster shapes (Murphy 2012). In particular, K-Means can be described as a special case of a clustering with a GMM using the EM Algorithm with a diagonal scalar covariance matrix, and a definite cluster assignment (Murphy 2012). Limitations of hierarchical methods are the missing mathematical optimization function (Murphy 2012) and the limited robustness, in case of changed data (Hastie et al. 2009). Nevertheless, K-Means often creates useful clustering results, and hierarchical algorithms are very useful for their explainability using dendrograms and since they do not need prior knowledge of the cluster amount (James et al. 2013). Therefore, two-stage approaches, in particular combining K-Means and Ward, are also common as recommended by Punj and Stewart (1983) or Li et al. (2011). Nonetheless, in these common methods, segment assignments are definite, and fuzzy assignments, as computed by GMMs, might be beneficial in business applications (Hiziroglu 2013).

In recent literature, the dominant approaches are still K-Means and agglomerative clustering with Ward in particular. In their review, Ernawati et al. (2021) discuss that clustering is the most applied method in customer segmentation, and within clustering, K-Means and its variants are the most commonly applied methods. In addition, Ghosal et al. (2020), in their review, present K-Means and hierarchical algorithms as the most common methods for a market analysis. Exemplary studies using clustering algorithms such as K-Means, Ward, agglomerative clustering, or modified K-Means models and as selected features recency, frequency, and monetary variables, often referred as RFM model, are given by Shihab et al. (2019); Abdulhafedh (2021); Aktaş et al. (2021); and Christy et al. (2021). In addition to an application of K-Means and Ward, Abdulhafedh (2021) utilizes a Principal Component Analysis and Aktaş et al. (2021) also include GMMs and Spectral Clustering in their benchmarking.

However, density-based clustering methods are rarely applied in customer segmentation due to their limitations, which are further discussed in following sections. Despite the low popularity and limitations, Hossain (2017) investigates DBSCAN, a common density-based clustering method, and states that it is capable of efficiently detecting outliers in customer segmentation. However, other authors who equally analyzed, compared, and benchmarked the clustering results of DBSCAN with K-Means or hierarchical approaches, often describe the results of DBSCAN as less useful, e.g., Banu (2022). The application of DBSCAN often results in the set-up of only one cluster if the clusters are too close to each other, e.g., in the analysis by Liço et al. (2021). Thus, Ghosal et al. (2020) describe banking as the main application domain of DBSCAN with tasks where unusual behavior must be separately detected.

Furthermore, in the commonly used approaches, key accounts are treated like every other customer without regarding their special position and business implications. In most approaches, all customers are analyzed together and no clear mechanism is set up to retrieve key accounts from the data set of customers.

Application of key accounts in customer segmentation and cluster analysis

The customer segment management within an organization is the customer support function regarding one specific type of product segment. The target of introducing customer segments is the specific alignment of activities of the marketing and sales departments to fulfill the customer segments' individual needs. The coordination of product

portfolios and customers is therefore a core competence of the marketing and sales departments. The key account management and customer segment management can be seen as a customer-centric coordination task within this organizational order (Homburg and Krohmer 2006).

The key account management is the support of a small amount of customers, so-called key accounts, and it is responsible for all products of the organization. Key accounts are attributed great importance due to their current or future purchasing behavior (Homburg and Krohmer 2006). There exist a variety of definitions and ideas to separate key accounts from other customers. The trivial approach is to select large customers with high current revenues since their cancelation of business relationships comes with a high operational risk for the organization (Sidow 2000). Another approach is the selection of complex and simultaneously high-revenue customers (Sidow 2000). The non-key accounts are defined to be more homogenous and not unique in their importance for the organization, which distinguishes them from key accounts (Homburg and Krohmer 2006).

However, there is no universally accepted or applied method and no clear mathematical definition to systematically and distinctly select key accounts. One common method for selecting key accounts is the ABC analysis using the revenues of the customers (Sidow 2000), (Ultsch and Lötsch 2015). While the ABC analysis will catch all high-revenue customers, it might underestimate the more complex customers or customers specialized within a certain product segment since not all product segments necessarily contribute an equal share to the total revenue. Therefore, if a data-centric analysis is conducted to segment customers or to find key accounts, the chosen approach must be capable of detecting customer segments as well as key accounts within the data, in addition to taking the organizational complexity of key accounts into consideration.

The following requirements for defining key accounts are formulated, considering the different key account definitions:

1. Key accounts should generate a high amount of revenue representing an important business relationship.
2. Key accounts should have a unique purchasing behavior which separates them from other customers. Therefore, they require a higher attention from sales.
3. Customers with lower revenue but with a highly important and complex business relation regarding one or multiple specific product segments should also be considered key accounts.
4. The amount of key accounts must be limited to a small number since the amount of key account managers in the sales department should be reasonably limited.

In particular requirement 2) corresponds with the idea of key accounts as data points deviating substantially from a normal assumption of customer behavior. Therefore, key accounts should be visible as anomalies or outliers in the data set of all customers. Thus, key accounts should be traceable using an anomaly detection approach. In addition, if key accounts are traceable as outliers within the customer base, they will also distort the data, and an exclusion should therefore result in a more robust statistical analysis of the customer segments.

If there exists within a company's customer base an amount of customers which agree to this key account definition, it is recommended to organizationally separate the customer support into a one-on-one relationship with a specifically commissioned key account manager while addressing the other customers over their assigned customer segment (Homburg and Krohmer 2006). Therefore, if customer segmentation is conducted, these key accounts must be separated from the other customers of the data set prior to the cluster analysis, since they are organizational supported outside of the built segments.

Methodology

Proposed two-step approach

Since it can be expected that there exist more low-revenue customers and that these customers individually are more similar to each other (Homburg and Krohmer 2006), these areas of low-revenue customers will be more densely populated in the data. This corresponds to the ABC analysis, where typically the low-revenue C-customers are expected to be more dense in the data (Ultsch and Lötsch 2015). If the area of key accounts is less dense than other areas, the key accounts will stand out when the data set is analyzed using a density-based anomaly detection or cluster analysis. A widespread algorithm for density-based cluster analysis is DBSCAN (Ester et al. 1996). In the following, DBSCAN's usefulness in the identification of less dense areas, and therefore key accounts, is analyzed. The approach could also apply LOF using an anomaly detection method (Breunig et al. 2000) or other density-based algorithms since the application of density-based methods for detecting key accounts leverages a core strength of this class of algorithms. This is highlighted by the outlier definition of LOF, which utilized an overall very similar approach to DBSCAN, by Pedregosa et al. (2011) which is that outliers are samples that have a substantially lower density than their neighbors. This definition can be applied to the entire field of density-based outlier detection and exactly reflects the property required for the outlier definition of key accounts as less homogenous and more unique data points.

After the identification of the less dense areas and the removal of key accounts, a GMM is applied and solved via an EM algorithm as a commonly used method in cluster analysis (Murphy 2012). A GMM seems favorable in contrast to other distance-based approaches, like K-Means or hierarchical clustering using a distance metric since GMMs utilize flexible covariance matrices and a fuzzy assignment. Regardless of which distance-based approach is chosen, the distance-based method's strengths over density-based approaches are leveraged in the second step since in the very dense areas containing non-key accounts, density-based methods could fail to separate customer segments in a beneficial manner. On the other hand, in approaches without prior filtering, distance-based approaches might be highly influenced by the key accounts due to their unique dissimilarity to all other customers and also to the other key accounts. This could be solved by, e.g., logarithmic scaling of the data or the application of a Principal Component Analysis, but then key accounts would no longer be weighted by their business importance and would not be separated as proposed.

Critical tasks in cluster analysis are the selection of a valid amount of clusters and the quality assessment of the set-up cluster assignment (Murphy 2012). Therefore, the method and improvement using the two-step approach need to be validated by these two criteria. In particular, a distinct clear-cut selection of the cluster amount is necessary if the GMM is applied in an automated as well as manual analysis.

The essence of the consideration of the proposed two-step approach for customer segmentation is to first exclude key accounts by density-based methods and second, to conduct a cluster analysis by distance-based methods. This idea leverages the definition of key accounts for an inclusion of domain-specific knowledge, creates a business-interpretation and business-explanatory friendly approach, and utilizes the density-based method's properties. Furthermore, this approach should result in a more robust statistical analysis of the data set.

The proposed approach using DBSCAN and GMMs, as described and conducted in the following sections, is given in Fig. 1.

Density-based anomaly detection

DBSCAN is a density-based cluster algorithm which separates clusters depending on how dense neighboring data points are. The core assumption is that the density of data points within a cluster is high, while the density between clusters is low. These areas of low density are then considered as noise. DBSCAN uses core data points which are defined by having at least *MinPts* data points, including itself, within a distance of *Eps*. Therefore, a cluster contains at least *MinPts* data points. Every data point within the distance *Eps* from a core data point belongs to the same cluster as the core data point. The core data points and all reachable data points build a cluster. The additional clusters are set up by core data points not reachable from the core data points of other clusters. Data points not reached by any cluster are considered as noise (Ester et al. 1996). DBSCAN is implemented as by Schubert et al. (2017) and is deterministic if the data are not permuted.

One challenge in the implementation of DBSCAN is the set-up of the parameters *MinPts* and *Eps*. Therefore, a k-dist-graph is proposed which assigns to each data point its distance to the k-th distant data point (Ester et al. 1996). If the distance to the k-th data point is smaller then *Eps* and $MinPts = k$, these data points build a cluster. If the k-dist-graph is sorted by decreasing k-th distance, a steep slope on the left becomes visible. These data points are the reachable data points, which are not core data points, and the defined

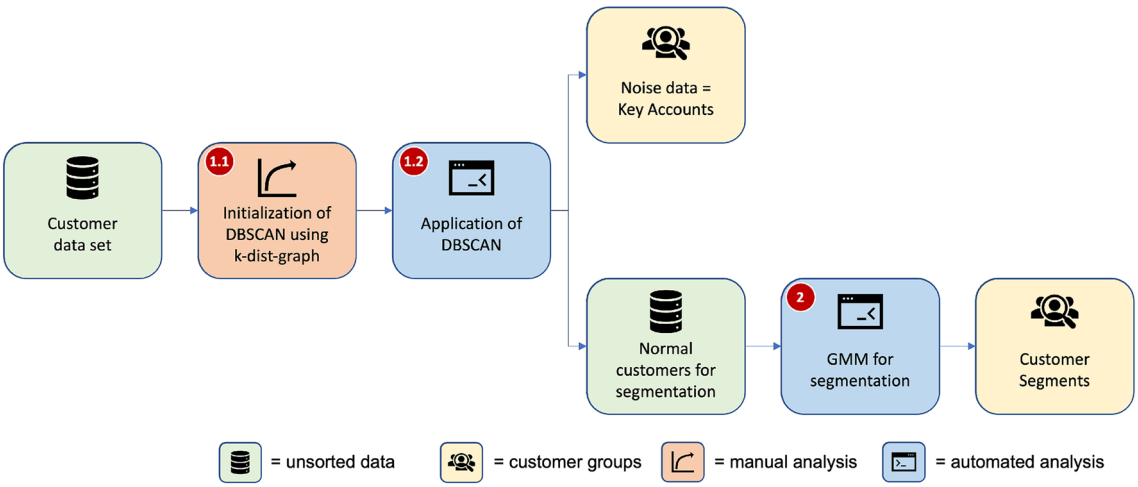


Fig. 1 Sketch of the proposed two-step approach with an application of DBSCAN and Gaussian Mixture Models

noise since the k-th distance is very high compared to other data points; therefore, they are unreachable from a core data point. It is proposed to use the bend after the slope from which the sorted k-dist-graph has a valley as a breaking point to find a valid *Eps* value (Ester et al. 1996). Also, it is recommended to use $MinPts = 4$ by Ester et al. (1996) since this value appears to be useful for many applications.

When the parameters are set up, it is necessary that the full data set and no subset or sample should be used. The parametrization using a subset might overestimate the applied value for *Eps* since the recommended k-th distance using the k-dist-graph is increased when less data populates the clusters. This might then result in outliers being assigned to clusters when the full data set is analyzed. Therefore, the k-dist-graph and DBSCAN should be evaluated using the full data set. This is complicated because the computational time complexity for setting up the k-dist-graph for N data points is $\mathcal{O}(N^2)$, since the k-nearest neighbor can be calculated in $\mathcal{O}(N)$ (Callahan and Kosaraju 1995), and the computational complexity for DBSCAN given by Schubert et al. (2017) is in the average case $\mathcal{O}(\log(N) \times N)$. Therefore, an analysis using the full data set requires an over-proportional run time. However, since the customer amount is comparatively limited in a B2B application, a customer segmentation must not be conducted very frequently in daily business activities, and no real-time requirements of the analysis have to be met, an analysis using the full customer data set is, in this described use case, still feasible. When a customer is added during daily business operations, it can easily be checked if its distance to any cluster exceeds *Eps* and can then be classified as key account. Otherwise, the customer can be classified by a following cluster analysis. Only if a large amount of customers are added, deleted, or changed is a renewal of the analysis necessary. Therefore, this limitation of DBSCAN does not affect the overall proposed approach.

Another common disadvantage of DBSCAN occurs during the clustering of data sets with varying densities of the clusters. In these scenarios, clusters with a higher density might be merged if a parameter *Eps* of a lower density cluster is applied or the lower density clusters might be considered noise if a too low *Eps* is applied. Since in the here-proposed application DBSCAN is only applied to detect the noise, not the clusters, this disadvantage can be neglected. This is shown by Breunig et al. (2000) in the application of a density-based anomaly detection for a benchmarking data set with differently dense clusters. Using LOF, anomalies are detected even when varying densities of the clusters are present. The outlined property is also applicable for DBSCAN if only the noise is relevant as long as the parameter *Eps* is set in a manner where outliers are not reachable from any cluster. If the k-dist-graph is set up for a data set with varying cluster densities, the different clusters are visible in the k-dist-graph as a rise between plateaus where a lower density cluster is visible from a higher, but within the same cluster overall stable, k-th distance of all data points of this cluster. Therefore, these plateaus do not affect the steeper slope of the outliers since they cannot be assigned to any cluster independently of the density and do not have a stable k-th distance. If the *Eps* corresponding to the cluster with the lowest density is selected as the corresponding parameter, this might result in a merging of clusters with higher density but will not affect the detected outliers since their k-th distance is still higher than *Eps*. Consequently, DBSCAN can be applied for the purpose of noise detection as proposed.

Gaussian mixture model for cluster analysis

A GMM assumes that each of N objects with a multivariate description x_i resulted from one multivariate normal distribution out of a total amount of K different multivariate normal distributions. Each distribution builds a cluster with its own cluster center μ_k, covariance matrix Σ_k , and weight within the overall data set π_k. The cluster parameters are combined in a parameter value θ_k. The distribution for the GMM for all K clusters is as follows (Murphy 2012):

$$p(x_i \mid \theta) = \sum_{k=1}^{K} \pi_k \, \mathcal{N}(x_i \mid \mu_k, \Sigma_k) \tag{1}$$

To compute the unknown cluster assignment of each object z_i, a responsibility r_{ik} that object i belongs to cluster k is formulated (Murphy 2012).

$$r_{ik} = \frac{p(z_i = k \mid \theta)\, p(x_i \mid z_i = k, \theta)}{\sum_{v=1}^{K} p(z_i = v \mid \theta)\, p(x_i \mid z_i = v, \theta)} \tag{2}$$

The responsibility is the assignment measure of object i to cluster k weighted by the sum of all assignment measures. Using the responsibility, the cluster analysis provides a fuzzy assignment of each customer to a cluster.

The estimation of the parameters is often archived by using the iterative EM algorithm. First, initial parameters are set for each cluster, including the assumed amount of clusters K. First step is the E-step (expectation) where the responsibility is updated using the initial parameters. The responsibility of iteration t uses the parameters of prior initialization θ^{t-1} (Murphy 2012).

$$r_{ik} = \frac{\pi_k \, p(x_i \mid \theta_k^{t-1})}{\sum_{v=1}^{K} \pi_v \, p(x_i \mid \theta_v^{t-1})} \tag{3}$$

After the expectation step, the M-step (maximization) is conducted by updating the parameters using the new responsibilities and cluster assignment (Murphy 2012).

$$\hat{\pi}_k = \frac{\sum_{i=1}^{N} r_{ik}}{N} \tag{4}$$

Using the updated responsibilities, the Maximum Likelihood (ML) estimator is optimized.

$$l(\mu_k, \Sigma_k) = \sum_{i=1}^{N} \sum_{k=1}^{K} r_{ik} \log(p(x_i \mid \theta_k)) \tag{5}$$

Using the ML-estimator, the cluster centers and corresponding covariance matrices are computed.

$$\hat{\mu}_k = \frac{\sum_{i=1}^{N} r_{ik} x_i}{\sum_{i=1}^{N} r_{ik}} \tag{6}$$

$$\hat{\Sigma}_k = \frac{\sum_{i=1}^{N} r_{ik} x_i x_i^T}{\sum_{i=1}^{N} r_{ik}} - \hat{\mu}_k \hat{\mu}_k^T \tag{7}$$

The cluster centers $\hat{\mu}_k$ are the mean of all objects' features weighted by the responsibilities. The covariance matrix $\hat{\Sigma}_k$ is the responsibility weighted deviation from the cluster center.

After the M-step, the E-step is repeated, iteratively calculating the responsibilities until a pre-defined termination criterion is reached. The final parameters describe the clusters' centers and covariance. All objects in the data set are assigned to the cluster with their respective highest responsibility.

Determination of cluster amount

Before selecting an applied GMM for measuring the results and setting up clusters, an important task is the determination of a valid model, i.e., the determination of a valid amount K of clusters. Therefore, a Bayesian approach is used to select the amount of clusters with the highest marginal likelihood (Murphy 2012). This approach has two disadvantages: the marginal likelihood is hard to calculate and a search through a high multitude of cluster amounts K is necessary.

The maximum likelihood estimator is using a pre-selected cluster amount K and is evaluated after estimating the parameters and conducting the cluster assignment (Murphy 2012). It is assumed that $\hat{L}(K)$ is the optimal value given a cluster amount K (Wit et al. 2012).

$$\hat{L}(K) = l(\theta, K) = \sum_{i=1}^{N} \log(p(x_i, \hat{z}_i \mid \theta)) \tag{8}$$

The amount of free parameters for a selected cluster amount K is $q(K)$. When using a fully free covariance matrix of an Gaussian Mixture Model, this results in the following term for an implementation of the GMM clustering (Pedregosa et al. 2011):

$$q(K) = K \times \left(\frac{D(D+1)}{2} + D + 1 \right) - 1 \tag{9}$$

Adding more clusters increases the likelihood of a model. Therefore, a penalty term is defined on the likelihood estimator for selecting a more valid cluster amount (Murphy 2012). A commonly used heuristic is the Akaike Information Criterion (AIC) defined as follows (Wit et al. 2012):

$$\text{AIC} = -2 \times \log\left(\hat{L}(K)\right) + 2 \times q(K) \tag{10}$$

A disadvantage of the AIC is that it does not penalize models with many clusters strictly enough. An information criterion which more strongly penalizes high amounts of clusters is given by the Bayesian Information Criterion (BIC) which results in more beneficial cluster amount recommendations in practical applications (Pedregosa et al. 2011). The BIC is defined as follows (Wit et al. 2012):

$$BIC = -2 \times \log\left(\hat{L}(K)\right) + q(K) \times \log(N) \tag{11}$$

To further penalize the BIC, (Biernacki et al. 2000) introduce the Integrated Completed Likelihood (ICL) criterion which additionally penalizes the BIC by adding the estimated mean entropy $E(K)$.

$$E(K) = -\sum_{k=1}^{K} \sum_{i=1}^{N} r_{ik} \log(r_{ik}) \geq 0 \tag{12}$$

The ICL criterion is then written as follows:

$$\text{ICL} = \text{BIC} + E(K) \tag{13}$$

A cluster amount K minimizing the applied information criterion should be selected. Using the BIC is recommended for multiple applications (Pedregosa et al. 2011) because a stricter penalty term results in a smaller amount of clusters. Under the assumption of Occam's Razor principle, this is beneficial (Murphy 2012).

Measurement of cluster quality

Without training data, the cluster shapes are commonly used for analyzing the quality of the proposed clusters (Igual and Segui 2017). An often-used score for measuring cluster quality is the silhouette coefficient (Pedregosa et al. 2011).

If C_k is a cluster from the set of all clusters C and the object $i \in C_k$ is assigned to cluster k, then $a(i)$, a support

variable which defines the silhouette coefficient, is the mean difference from object i to all other objects $u \in C_k$ with $u \neq i$ using the selected metric (Rousseeuw 1986).

$$a(i) = \frac{1}{\|C_k\|} \sum_{i \neq u} \Delta(x_i, x_u) \tag{14}$$

The support variable $b(i)$ is defined as the smallest mean dissimilarity of object i to the other clusters C_v with $v \neq k$ (Rousseeuw 1986).

$$b(i) = \min_{C_v \neq C_k} \Delta(x_i, C_v) \tag{15}$$

The silhouette coefficient is then given as follows (Rousseeuw 1986):

$$s(i) = \frac{b(i) - a(i)}{max(a(i), b(i))} \tag{16}$$

The silhouette coefficient is bounded by $-1 \leq s(i) \leq 1$. A score of 1 is interpreted as a perfect cluster quality, whereas the score of -1 is the least beneficial cluster quality. 0 acts as a neutral evaluation (Igual and Segui 2017). If a cluster contains only one object, the score is set to 0 (Rousseeuw 1986).

It is proposed to use the mean of all silhouette coefficients as a measure of the quality of the cluster analysis and also as a beneficial metric for interpretation and validation of the cluster analysis (Rousseeuw 1986).

One disadvantage of the silhouette coefficient is its preference for spherical clusters (Rousseeuw 1986), but, for a better understanding in real world applications of cluster analysis results, spherical clusters are often preferred.

Use-case: customer segmentation for a B2B wholesaler of food & household supplies

Exclusion of key accounts using DBSCAN

The proposed method is applied to a data set of the customers' revenue per product segment of a Portuguese wholesaler for food and household supplies (Abreu 2011) which is in this paper examined using scikit-learn (Pedregosa et al. 2011) as implementation of the Gaussian Mixture Model's solver with maximum likelihood estimations.

Other authors have already used the data set to research regional differences within the customer groups (Natesh and Shobha Rani 2018) and conducted customer segmentations (Baudry et al. 2012) which makes it suitable as a test object for our proposed method. Notably, the data set was already analyzed applying a GMM by Baudry et al. (2012) using different information criteria for cluster analyses and applying different amounts of customer groups K within the data set but without the removal of key accounts. Therefore, the research by Baudry et al. (2012) is used for a comparison of results. Also, Baudry et al. (2012) were unable to define a valid cluster amount applied in the analysis using the more commonly used BIC and ICL. Therefore, the analysis will be compared to examine how the exclusion of key accounts can benefit the selection of the cluster amount using information criteria.

The data set contains 440 Portuguese customers of said wholesaler with revenues in six product segments of food and household supplies. For each of the customers, the individual revenue (in monetary units) per product segment, its region, and its industry sector, either hotel/restaurant/cafe or retail, is given. 77 customers are from the region Lisbon, 47 are from the region Porto, and 316 are considered from "other" regions. Regarding the industry sector, 142 customers are in retail, 298 customers are hotels/restaurants/cafes.

The focus of the analysis is the revenue per product segment since within customer segment management the customers are separated based on purchasing behaviors in different product segments (Homburg and Krohmer 2006). The descriptive statistics of the product segments is given in Table 1.

The existence of high-revenue customers is visible within the descriptive statistics. The maximum revenues are more than 5 standard deviations distant from the mean values. Also, since the mean is low compared to the maximum values with simultaneous high standard deviation, it can be hypothesized that low-density areas of key accounts exist. Therefore, the k-dist-graph (Ester et al. 1996) becomes a useful tool in detecting the key accounts. As initially proposed by Ester et al. (1996), the analysis is conducted using

Table 1 Revenue per product segment

Product segment	Min revenue	Max revenue	Mean revenue	Std. revenue
Fresh	3.00	112,151.00	12,000.30	12,647.33
Milk	55.00	73,498.00	5796.27	7380.38
Grocery	3.00	92,780.00	7951.28	9503.16
Frozen	25.00	60,869.00	3071.93	4854.67
Detergents/paper	3.00	40,827.00	2881.49	4767.85
Delicacies	3.00	47,943.00	1524.87	2820.11

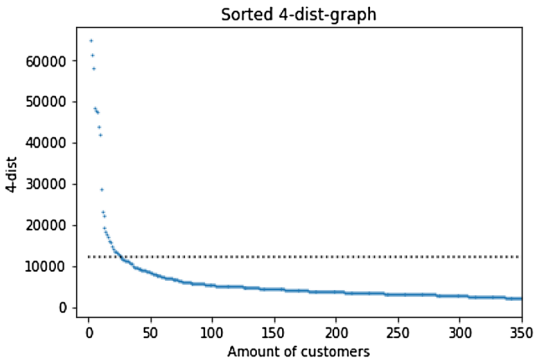


Fig. 2 Applied 4-dist-graph for the wholesaler customer data. The proposed parameter of $Eps = 12,000$ is added as the dotted straight line. Data points below the *Eps* value are core data points. Data points above are either reachable data points or noise

$k = 4$ and an *Eps* value is selected, where the k-dist-graph flattens. The k-dist-graph is given in Fig. 2.

Using the proposed heuristic by Ester et al. (1996) in Fig. 2, customers with a distance $>$ 12,000 to the nearest cluster should be considered key accounts since the curve flattens after this value. Also, since the C-customers are more dense, the ABC analysis suggests (Ultsch and Lötsch 2015) that the customers within the noise of DBSCAN might be complex and high-revenue key accounts. Therefore, DBSCAN is applied using $Eps = 12{,}000$ and $MinPts = 4$ classifying 16 customers as noise. The descriptive statistics of the customers within noise is given in Table 2.

All 16 identified customers are either part of the 5% revenue strongest customers or are part of the top three customers in revenue per product segment. All maximum revenues per product segments are from the group of customers within noise. Of the 22 overall 5% revenue strongest customers, 14 are considered noise. Also, as shown in Table 2, the standard deviation within the noise customers is quite high. This suggests that while the customers are from the set of customers with the highest revenue overall, they have a unique purchasing behavior and do not demand each product segment equally. This is the result not only of a high spending but also of a very targeted and complex purchasing behavior.

Thus, it is proposed to define these customers within the less dense area of noise as key accounts since the customers within the noise fulfill all requirements defined in the "Application of key accounts in customer segmentation and cluster analysis" section. Note that this key account definition used in the study is more focused on complexity of customers than just revenue (Sidow 2000) but aligns with the less dense A-customers of an ABC analysis. This is manifested by the fact that not all high-revenue customers are defined as key accounts, especially if the purchasing behavior of a high-revenue customer is very similar or comparable to the purchasing behavior of smaller customers. Since the purchasing behavior is less complex in these cases, there is no special need for consulting by the sales department, and the customer can be treated organizationally more like a normal customer. Vice versa, 2 customers with smaller revenues compared to the 22 overall 5% revenue strongest customers are defined as key accounts since their purchasing behavior is more complex and specialized in certain product segments.

Since key accounts are supported outside of the regular customer segments, they are not used for the following customer segmentation. The descriptive statistics of the customer data without key accounts is given in Table 3.

Table 2 Revenue per product segment of customers within noise

Product segment	Min revenue	Max revenue	Mean revenue	Std. revenue
Fresh	85.00	112,151.00	36,496.8	27,726.3
Milk	1266.00	73,498.00	24,647.6	21,399.7
Grocery	2062.00	92,780.00	30,881.9	25,044.3
Frozen	36.00	60,869.00	13,598.6	16,349.9
Detergents/paper	71.00	40,827.00	12,027.0	14,047.8
Delicacies	903.00	47,943.00	8165.1	11,463.0

Table 3 Revenue per product segment without key accounts

Product segment	Min revenue	Max revenue	Mean revenue	Std. revenue
Fresh	3.00	56,159.00	11,075.90	10,635.10
Milk	55.00	29,892.00	5084.89	5019.91
Grocery	3.00	39,694.00	7085.97	7017.52
Frozen	25.00	18,711.00	2674.70	3158.29
Detergents/paper	3.00	19,410,00	2536.38	3579.74
Delicacies	3.00	7844.00	1274.30	1244.66

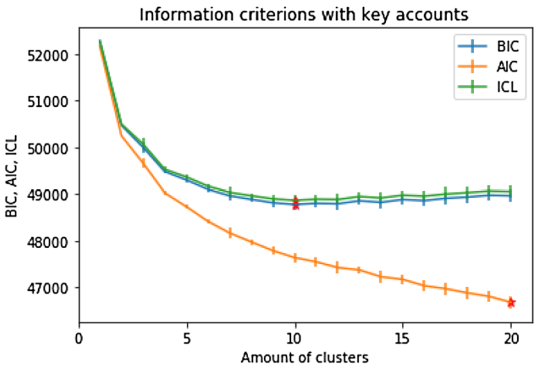


Fig. 3 Information criteria for the wholesaler customers with key accounts using 30 cluster assignments with 10 initializations each of the EM algorithm. The 1-σ standard deviation is visualized as an error bar. The cluster amounts minimizing the criteria are marked in red. (Color figure online)

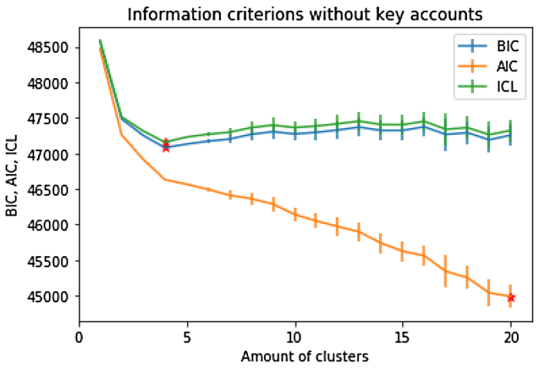


Fig. 4 Information criteria for the wholesaler customers without key accounts using 30 cluster assignments with 10 initializations each of the EM algorithm. The 1-σ standard deviation is visualized as an error bar. The cluster amounts minimizing the criteria are marked in red. (Color figure online)

An overview of the customers without key accounts in Table 3 shows that the customers' purchasing data are more dense since DBSCAN finds only one cluster. Also, the customers are more homogeneous and not special regarding their relationship to the wholesaler in one product segment, which matches the business science description (Homburg and Krohmer 2006). Within business operations, customer segments should be defined as convex clusters and avoid spiral or crescent formed cluster shapes, so if a customer has a revenue in each product segment between two customers of the same clusters, it should also belong to the same cluster. Therefore, even if DBSCAN results in clusters, the non-convex cluster shapes of DBSCAN would be less useful.

It can be concluded that the proposed approach using DBSCAN is capable of identifying customers which can be considered key accounts regarding complexity and revenue size. All identified customers within the noise can be defined as key accounts either by a revenue measure as well as by complexity of purchasing and therefore, it is concluded that our approach is aligned with the business implications and practical applications of key accounts.

Determination of cluster amount

The next task is the selection of the right cluster amount. To analyze the information criteria, the EM algorithm is initialized multiple times and the analysis is conducted over multiple simulations to create more meaningful mean values and error bars. Note that the standard deviation becomes very small for smaller customer amounts.

If the analysis is conducted with all customers, the information criteria give no clear recommendation. Figure 3 would recommend a high amount of customer segments (≥ 10) but the analysis lacks a distinct minimum for one specific cluster amount, and the curves of the ICL and BIC have a plateau without a distinct clear-cut selection. The curves of the information criteria are similar to the results of the prior analysis by Baudry et al. (2012) which also computed a recommended cluster amount of $K = 10$. Since lower amounts of clusters are often preferred and no clear-cut selection is possible, Baudry et al. (2012) introduced a new criterion which found mixed recommendations ranging from 3 to 5 recommended customer segments.

Once the key accounts are separated from the normal customers within the data set, a more distinct analysis of the cluster amount minimizing the information criteria values is possible. The information criteria of the customer data set without key accounts are evaluated in Fig. 4.

The analysis in Fig. 4 shows a distinct minimum value for the BIC and ICL at 4 clusters. Only the AIC does not recommend a valid cluster amount since it penalizes the increase in clusters too little but shows a sharp reduction of the curve's slope at 4 clusters. Not only is the information density increased, expressed by lower criterion values, but the analysis also becomes easier to interpret and the plots more distinct. Thus, no additional information criterion is necessary as the commonly used criteria perform well. This result suggests the usefulness of delimiting key accounts before segmentation based on their special position and different purchasing behavior.

In the comparative study, the selection of the cluster amount is more complicated (Baudry et al. 2012). If key accounts are not separated, the preferred cluster amount by Baudry et al. (2012) using additional information criteria results in three clusters since it offers the most distinct purchasing behavior while also providing a good allocation of the channels. This recommendation is therefore used as comparison for the clustering results.

In a later publication, Baudry et al. (2015) used a Principal Component Analysis to reduce the dimensionality of

the data set and to compute more meaningful results. The cluster amount then recommended is $K = 5$ using the BIC with key accounts included, which is similar to the here-proposed cluster amount of 4 customer segments plus 1 key account group. The exclusion of key accounts is an application of domain knowledge, i.e., the widely used application of key accounts in marketing sciences, and might improve an analysis from a business domain's perspective to become more meaningful and create a greater explanatory value than the conducted Principal Component Analysis. Additionally, within the solution of Baudry et al. (2015) the anomalously behaving key accounts might still distort the segments and might result in less beneficial business decisions.

To summarize, the proposed two-step approach excluding key accounts from the data is capable of increasing the usability of the model selection and results in more distinct recommendations using information criteria.

Discussion of clustering results

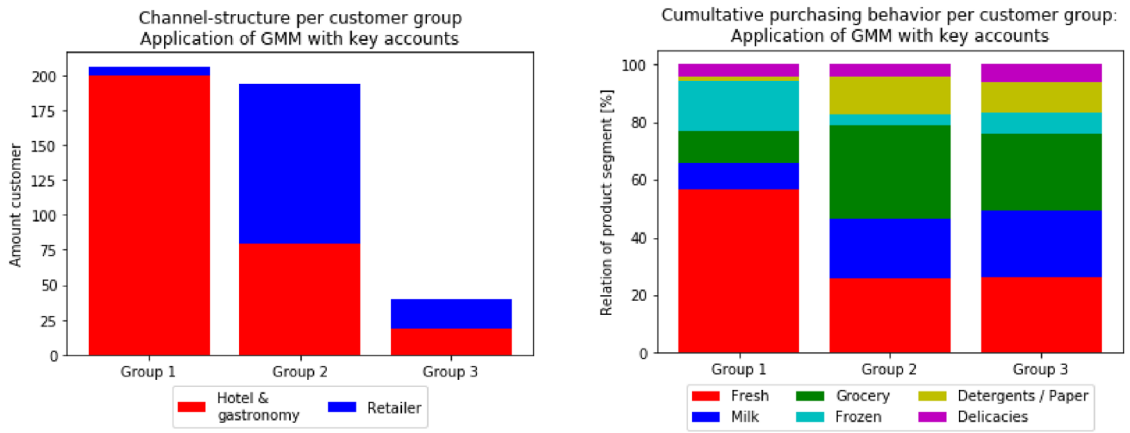
The 3-cluster solution suggested by Baudry et al. (2012) uses a key account cluster (group 3), a retail-heavy cluster (group 2), and a hotel/gastronomy cluster (group 1), see Fig. 5a. Within the 3-cluster solution, only group 1 is able to distinctly separate the customer channels retail and hotel/gastronomy. Since the data is more distorted due to the included key accounts, the usage of a full covariance matrix is beneficial for the approach.

Without key accounts, the analysis using four clusters (and one key account group) in Fig. 6a shows the A and B customers, as defined by the ABC analysis (Ultsch and Lötsch 2015), of hotels/gastronomy (group 2, B-customers with lower revenue per customer and group 3, A customers with higher revenue per customer). Moreover, the analysis separates the A and B customers (group 4) of retail with a very distinct mix of product purchasing behaviors in Fig. 6b. Additionally, the analysis shows a C-customer segment (group 1), very likely representing small cafes, shops, and takeaways, with a wide variety of demands over all product segments. One can note that a spherical covariance matrix was used in this case to increase the quality of results since the data are less distorted without the key accounts. Also, a spherical covariance matrix makes the results more comparable to results archived by applications of K-Means since K-Means is a special case of the EM algorithm with definite cluster assignment and spherical clusters (Murphy 2012).

An overview of the amount and cumulative revenue per cluster in the 3-cluster solution with key accounts and the 4+1-cluster solution with separated key accounts is given in Table 4.

These results nicely highlight the strength of our proposed approach. In contrast to Baudry et al. (2012) our results contain additional information on A and B customers of retail and hotels/gastronomy by separating them into distinct groups, while also separating the miscellaneous category containing small cafes and takeaways from the bigger hotels/gastronomy. Channels within the 4-cluster approach leaving out key accounts are overall more distinctly separated. Although Baudry et al. (2012) claim that a 5-cluster approach (4 clusters + 1 key account cluster) is less usable and favor a 3-cluster approach, further insights are gained by using our proposed method with 4 + 1 clusters.

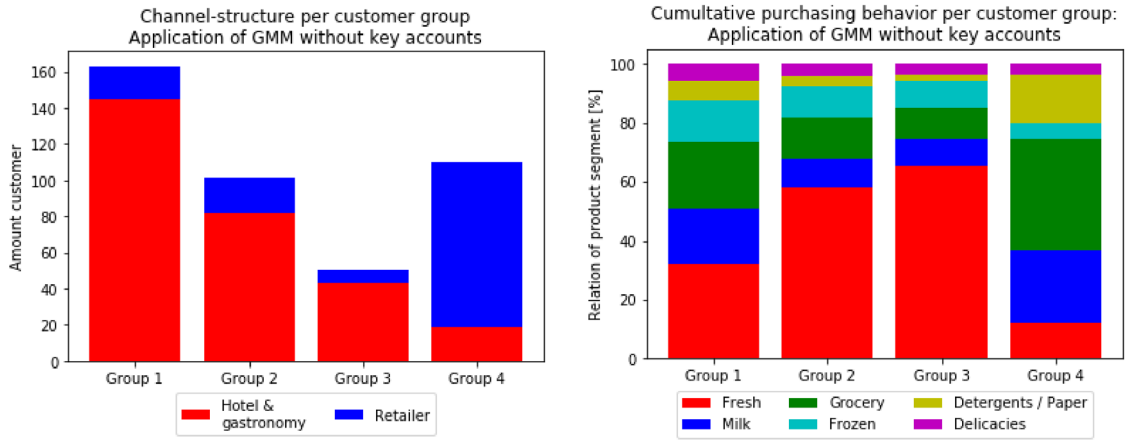
For a cluster quality assessment, the silhouette coefficients (using spherical covariance matrices) in Table 5 show no major difference between analyzing 4 clusters without key accounts ($s = 0.311$) and the recommended



(a) Channel structure per segment.

(b) Purchasing behavior per segment.

Fig. 5 Clustering using Gaussian Mixture Model (GMM) with included key accounts for $K = 3$ groups using 1000 initializations and full covariance matrices



(a) Channel structure per segment. (b) Purchasing behavior per segment.

Fig. 6 Clustering using Gaussian Mixture Model (GMM) with excluded key accounts for $K = 4$ groups using 1000 initializations and spherical covariance matrices

Table 4 Customer amounts and cumulative revenue per cluster

	Excluded key accounts		Included key accounts	
	Customer amount	Cumulative revenue	Customer amount	Cumulative revenue
Group 1	163	2,473,679	210	4,765,025
Group 2	101	2,818,211	189	6,325,530
Group 3	50	2,575,756	41	3,528,945
Group 4	110	4,738,782	–	–
Key accounts	16	2,013,072	Compare with group 3	

Table 5 Silhouette coefficients for small cluster amounts

Cluster amount	Silhouette coefficients	
	Excluded key accounts	Included key accounts
3 Cluster	0.376	0.255
4 Cluster	0.311[a]	0.284
5 Cluster	0.295	0.286[a]

[a]Recommended cluster amounts

Clustering using Gaussian Mixture Model with 1000 initializations and spherical covariance matrices

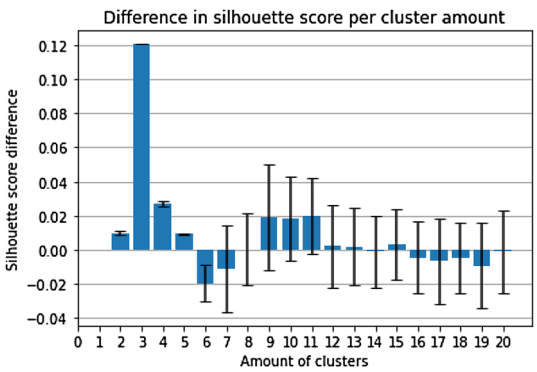


Fig. 7 Differences of the silhouette coefficients for different cluster amounts using 30 cluster assignments with 10 initializations each of the EM algorithm and spherical covariance matrices. A positive value indicates a higher coefficient if key accounts are excluded. The 1-σ standard deviation is visualized as an error bar

cluster amount by Baudry et al. (2012) with key accounts ($s = 0.286$). Nevertheless, the exclusion of key accounts results in a slight increase of the silhouette coefficients used as quality-assessment criteria. This is the case since the less dense and divergent key account areas in the data set distort the spherical form of the clusters. This effect is more significant for smaller cluster amounts since the key accounts are then mixed with smaller customers. Overall,

quality assessment of clusters is often very case specific and, more importantly, the higher linkage between channels and customer segments indicate results with greater explanatory value.

The differences in the silhouette coefficients between an analysis with included and excluded key accounts can be analyzed using different amounts of clusters. The analysis is given in Fig. 7. The exclusion of key accounts is, when using the silhouette coefficient as quality assessment, useful for smaller amount of clusters. For cluster amounts higher than five clusters, no significant effect is measured since the key accounts can be grouped into smaller clusters, resulting in less distortion of the larger clusters. For cluster amounts of five clusters and fewer, the results are highly significant and indicate a clear improvement of the cluster quality if key accounts are excluded from the analysis. Since in the application of cluster analysis in B2B a smaller cluster amount is assumed to be beneficial, as discussed in the section "Application of key accounts in customer segmentation and cluster analysis", the exclusion of key accounts improves the result in these cases in particular. Overall it can be concluded that the proposed method is therefore beneficial for cluster analysis using smaller cluster amounts but does not change the cluster quality significantly if used for larger cluster amounts.

The results indicate that the separation of key accounts benefits the model identification and also provides a more insightful analysis of the customer groups, i.e., since it provides a better link between customer segments and customer channels. Thus, the proposed approach provides marketing departments with a method to define customer segments and key accounts through a data-centric measure and enables a refocus from pure revenue-based approaches to purchasing complexity-driven approaches. Also, the proposed approach enables a more distinct analysis and therefore provides an easier implementation of clustering for customer segmentation in practical applications.

Conclusion

The proposed two-step approach using a density-based outlier detection and then a GMM for clustering customer segments is able to leverage multiple benefits as shown in the case study of a B2B customer segmentation of a Portuguese wholesaler.

First, key accounts are separated as recommended from the remaining customers applying a common definition of key accounts as complex and important customers, while using all available customer revenue data. The separated key accounts are unique in their complexity, and the density-based approach is aligned in its consideration with the commonly used ABC analysis.

Second, the removal of key accounts enables an easier selection of the necessary parameters in setting up the cluster algorithms. The commonly used BIC criterion results in more meaningful recommendations after the removal of key accounts and enables an easier decision regarding cluster amount compared to other studies without out this approach.

Third, the set-up customer segments are also more meaningful and useful in differentiating the customers types. All applied quality criteria of the cluster analysis indicate slightly better results. This might not be surprising since outliers are well known to distort results. Additionally, by applying the key account definition as a business rationale, outliers are not simply ignored but explicitly included in the analysis and handled separately.

The proposed approach provides a more streamlined analysis of customer segments and enables a mathematical model to define key accounts and customer segments based on complexity, as well as revenue to improve management decisions in marketing and sales. Furthermore, the conducted analysis and in particular the improved results of the customer segments show the importance of adding domain-specific knowledge to an analysis, in this case key accounts as a commonly used concept of marketing management. A pure data-centric approach might be limited, and this contribution provides another example that data analysis is often improved by combining the knowledge of multiple domains and the organizational data owners.

In conclusion, the proposed approach provides, besides a definition for selecting key accounts, a solution for the common limitations of cluster analysis for customer segmentation. Most notably, the method results in less distortion of segments by outliers, the usage of a fuzzy cluster assignment instead of a definite one, a more distinct and easier set-up of the valid cluster amount, and an increased explanatory value of the key account segment. A limitation of the proposed approach is that the selection of the covariance matrix type and the set-up of DBSCAN are still conducted manually.

Further research should be conducted to compare different anomaly detection methods besides DBSCAN for identifying key accounts and to systematically test the usefulness of the approach for different use cases, industrial sectors, and data sets.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflict of interest The author states that there is no conflict of interest.

References

Abdulhafedh, A. 2021. Incorporating K-means, Hierarchical Clustering and PCA in Customer Segmentation. *Journal of City and Development* 3 (1): 12–30.

Abreu, N. 2011. Analise do perfil do cliente Recheio e desenvolvimento de um sistema promocional. Repository ISCTE-IUL. Lisbon: Instituto Universito de Lisboa. http://hdl.handle.net/10071/4097).

Aggarwal, C. 2013. *Outlier Analysis*. New York: Springer.

Aktaş, A.A., O. Tunalı, and A.T. Bayrak. 2021. Comparative Unsupervised Clustering Approaches for Customer Segmentation. In *2021 2nd International Conference on Computing and Data Science (CDS)*.

Banu, T. 2022. Customer Segmentation with Machine Learning for Online Retail Industry. *The European Journal of Social & Behavioural Sciences* 31 (2): 111–136.

Baudry, J., M. Cardoso, G. Celeux, M. Amorim, and A. Ferreira. 2012. Enhancing the selection of a model-based clustering with external qualitative variables. arXiv Preprint. arXiv:1211.0437.

Baudry, J., M. Cardoso, G. Celeux, M. Amorim, and A. Ferreira. 2015. Enhancing the selection of a model-based clustering with external qualitative variables. *Advances in Data Analysis and Classification* 9 (2): 177–196.

Berkhin, P. 2006. A Survey of Clustering Data Mining Techniques. In *Grouping Multidimensional Data*, ed. J. Kogan, C. Nicholas, & M. Teboulle. Berlin: Springer.

Biernacki, C., G. Celeux, and G. Govaert. 2000. Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (7): 719–725.

Breunig, M., H.-P. Kriegel, R. Ng, and J. Sander. 2000. LOF: Identifying Density-Based Local Outliers. *ACM SIGMOD Record* 29 (2): 93–104.

Brudvig, S., M. Brusco, and J. Cradit. 2019. Joint selection of variables and clusters: Recovering the underlying structure of marketing data. *Journal of Marketing Analytics* 7 (1): 1–12.

Callahan, P.B., and S.R. Kosaraju. 1995. A Decomposition of Multidimensional Point Sets with Applications to K-Nearest-Neighbors and n-Body Potential Fields. *Journal of the ACM* 42 (1): 67–90.

Chandola, V., A. Banerjee, and V. Kumar. 2009. Anomaly detection: A survey. *ACM Computing Surveys* 41 (3), Article 15.

Christy, A.J., A. Umamakeswari, L. Priyatharsini, and A. Neyaa. 2021. RFM ranking—An effective approach to customer segmentation. *Journal of King Saud University - Computer and Information Sciences* 33 (10): 1251–1257.

Ernawati, E., S.S.K. Baharin, and F. Kasmin. 2021. A review of data mining methods in RFM-based customer segmentation. *Journal of Physics: Conference Series* 1869 (1): 012085.

Ester, M., H.-P. Kriegel, J. Sander, and X. Xu. 1996. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, pp. 226–231. Portland: AAAI Press.

Ghosal, A., A. Nandy, A.K. Das, S. Goswami, and M. Panday. 2020. A Short Review on Different Clustering Techniques and Their Applications. In *Emerging Technology in Modelling and Graphics*, ed. J. K. Mandal and D. Bhattacharya, 69–83. Singapore: Springer.

Hastie, T., R. Tibshirani, and J. Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer.

Hiziroglu, A. 2013. A neuro-fuzzy two-stage clustering approach to customer segmentation. *Journal of Marketing Analytics* 1 (4): 202–221.

Homburg, C., and H. Krohmer. 2006. *Marketingmanagement: Strategie - Instrumente - Umsetzung - Unternehmensführung*. Wiesbaden: Betriebswirtschaftlicher Verlag Dr. Th. Gabler.

Hossain, A.S. 2017. Customer segmentation using centroid based and density based clustering algorithms. In *2017 3rd International Conference on Electrical Information and Communication Technology (EICT)*, pp. 1–6.

Huber, P., and E. Ronchetti. 2009. *Robust Statistics*. Hoboken: Wiley.

Igual, L., and S. Segui. 2017. *Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications*. Cham: Springer.

James, G., D. Witten, T. Hastie, and R. Tibshirani. 2013. *An Introduction to Statistical Learning*. New York: Springer.

Jensen, O. 2001. *Key-Account-Management: Gestaltung - Determinanten - Erfolgsauswirkungen*. Wiesbaden: Deutscher Universitäts-Verlag.

Li, D.-C., W.-L. Dai, and W.-T. Tseng. 2011. A two-stage clustering method to analyze customer characteristics to build discriminative customer management: A case of textile manufacturing business. *Expert Systems with Applications* 38 (6): 7186–7191.

Liço, L., I. Enesi, and B. Çiço. 2021. Analyzing Performance of Clustering Algorithms on a Real Retail Dataset. In *2021 International Conference on Information Technologies (InfoTech)*, pp. 1–6.

Mehrotra, K., C. Mohan, and H. Huang. 2017. *Anomaly Detection—Principles and Algorithms*. Cham: Springer.

Murphy, K. 2012. *Machine Learning: A Probabilistic Perspective*. Cambridge: Massachusetts Institute of Technology.

Natesh, T., and N. Shobha Rani. 2018. Customer Puzzled Behavioral Analysis—A Step Towards Valuing Customer's Interests. *International Journal of Mechanical Engineering and Technology* 9 (7): 365–374.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12: 2825–2830.

Punj, G., and D.W. Stewart. 1983. Cluster Analysis in Marketing Research: Review and Suggestions for Application. *Journal of Marketing Research* 20 (2): 134–148.

Rousseeuw, P. 1986. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 20: 53–65.

Schubert, E., J. Sander, M. Ester, H.-P. Kriegel, and X. Xu. 2017. DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN. *ACM Transactions on Database Systems (TODS)* 42 (3), 1–21, Article 19.

Shihab, S.H., S. Afroge, and S.Z. Afroge. 2019. RFM Based Market Segmentation Approach Using Advanced K-means and Agglomerative Clustering: A Comparative Study. In *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, pp. 1–4.

Sidow, H. 2000. *Key Account Management. Wettbewerbsvorteile durch kundenbezogene Strategien*. Landsberg/Lech: Verlag Moderne Industrie.

Ultsch, A., and J. Lötsch. 2015. Computed ABC Analysis for Rational Selection of Most Informative Variables in Multivariate Data. *PLoS ONE* 10 (6): e0129767.

Wit, E., E. van den Heuvel, and J. Romeijn. 2012. All models are wrong: An introduction to model uncertainty. *Statistica Neerlandica* 66 (3): 217–236.