**COUSTOMER SEGMENTATION AND PERSONALIZATION**

**A Project Report**

**submitted in partial fulfillment of the requirements**

**of**

**……………. Track Name ……**

by

**Name of Student, G CHANDRU**

**Email id, chandruotk@gmail.com**

**NM Id,** 34E32607EA40F2AB6A6465AD8E3A0ECC

Under the Guidance of

**(P. Raja, Master Trainer, Edu net Foundation)**

**ACKNOWLEDGEMENT**

I’m glad that to receive an amazing opportunity to endure my knowledge and skills.

First and foremost, I want to extend my sincere gratitude to my supervisor **P. Raja, Master Traine**r, who has been an exceptional mentor and an invaluable adviser throughout this journey. His guidance, encouragement, and constructive feedback have ignited numerous innovative ideas and have been pivotal in the successful completion of this project. The trust he has shown in my abilities has been incredibly motivating, and it has been a privilege to work with him over the past year. He has always provided support not only for my project but also for various elements of the program. His wisdom and teachings have significantly contributed to my development as a capable and responsible professional.

I am also profoundly thankful for the resources and support extended by my college **MOHAMED SATHAK ENGINEERRING COLLEGE**. Because of their support and providing necessary equipment also providing take **NAAN MUDALVAN** course.

*ABSTRACT*

This paper explores the application of Artificial Intelligence (AI) techniques for customer segmentation and personalization to enhance customer engagement and optimize marketing strategies. By leveraging customer data, including preferences, behaviors, and purchase history, AI-driven models enable the segmentation of customers into distinct groups with shared characteristics. These segments allow for a more nuanced understanding of customer needs and enable personalized recommendations and targeted marketing campaigns, improving both customer satisfaction and business outcomes. Key AI techniques, such as machine learning algorithms, clustering, and recommendation systems, are discussed in the context of their role in creating dynamic and personalized customer experiences. The findings indicate that adopting AI for customer segmentation and personalization can lead to significant increases in customer engagement, conversion rates, and long-term brand loyalty.

**TABLE OF CONTENTS**

Abstract 3

List of Figures 4

List of Tables5

**Chapter 1. Introduction**

1.1Problem Statement

1.2Motivation

1.3Objectives

1.4. Scope of the Project

**Chapter 2. Literature Survey**

2.1 Literature overview

2.2 Indepth customer segmentation methods

2.3 Overview customer targeting use cases

**Chapter 3. Proposed Methodology**

3.1 System Design

3.2Overview and examples of the interplay between customer representation and analysis for customer targeting use cases

**Chapter 4. Implementation and Results**

* 1. Implementation

4.2 Results of Customer Segmentation and Personalization

**Chapter 5. Discussion and Conclusion**

5.1 Analysis and discussion

5.2 Analysis of feature selection methods

5.3 Analysis of segmentation methods

5.4 Challenges and Best Practices

**6 Conclusion and future research**

**7 References**

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **Figure No** | **Name of the Figure** | Page No. |
|  | Flow chart of the literature research process. | 11 |
|  | Process of customer targeting based on behavioral information gathered from data. | 13 |
|  | Distribution of surveyed publications from 2000 until 2022. | 14 |
|  | Distribution of the surveyed feature selection methods over the year. | 14 |
|  | Distribution of the surveyed clustering methods over the years of publication. | 15 |
|  | Distribution of the segmentation methods used evaluation methods. | 23 |

**LIST OF TABLES**

|  |  |  |
| --- | --- | --- |
| **Table No** | **Name of the table** | **Page No.** |
| Table 1 | Use cases and utilized segmentation approaches for customer targeting of commerce business from the survey literature. | 17 |
| Table 2 | Segmentation and used feature selection methods with corresponding references of the survey literature. | 18 |
| Table 3 | Data dimensionality that the feature selection methods and clustering approaches handle in the experiments of the surveyed literature | 22 |
| Table 4 | Literature with title, author and date which are the object of investigation sorted by year of publication/acceptance | 42-46 |

**CHAPTER 1**

**Introduction**

**“**As the Internet emerges as a new marketing channel, analyzing and understanding the needs and expectations of their online users or customers are considered as prerequisites to activate the consumer-oriented electronic commerce. Thus, the mass marketing strategy cannot satisfy the needs and expectations of online customers. On the other hand, it is easier to extract knowledge out of the shopping process under the Internet environment. Market segmentation is one of the ways in which such knowledge can be represented and make it new business opportunities.”

The rise of Artificial Intelligence (AI) has opened up new avenues in the field of customer segmentation and personalization, offering businesses more granular insights into consumer behavior and preferences. Traditional marketing approaches often rely on broad demographic categories and basic customer profiles, leading to a one-size-fits-all approach in marketing. AI-driven techniques, however, allow for more sophisticated segmentation by analyzing patterns in customer data, including browsing behavior, purchase history, and interactions. This shift enables businesses to move beyond static segments and deliver highly personalized recommendations and targeted campaigns. With consumers increasingly expecting relevant and timely interactions, AI-based personalization has become essential for enhancing customer experience, improving conversion rates, and fostering brand loyalty.

**1.1. Problem Statement**

In today’s competitive market, businesses struggle with delivering relevant marketing messages to individual customers. Traditional segmentation methods often fail to capture the full spectrum of customer behaviors, leading to ineffective marketing strategies and missed revenue opportunities. Without accurate and actionable insights, companies risk alienating customers with irrelevant or untimely messages. The challenge is to leverage vast and diverse customer data to create dynamic segments and develop personalized marketing strategies that resonate with each customer. This project aims to address these challenges by applying AI techniques to improve the accuracy of customer segmentation and personalization efforts, enabling businesses to target their audiences more effectively and efficiently.

**1.2 Motivation**

The motivation behind using AI for customer segmentation and personalization stems from the evolving needs of businesses to engage customers on a more individual level in today’s data-driven landscape. Traditional marketing approaches often struggle to address the diverse and unique preferences of each customer, resulting in generalized campaigns that lack relevance. In a market where customers expect tailored experiences, generic approaches can lead to low engagement, reduced customer satisfaction, and missed revenue opportunities.

With the growing volume of customer data from online interactions, purchase histories, and preferences, businesses have a significant opportunity to understand and anticipate customer needs. AI techniques enable organizations to harness this data effectively by identifying meaningful customer segments and delivering personalized experiences that resonate with each segment.

Furthermore, personalization has been shown to positively impact key business metrics such as customer engagement, conversion rates, and brand loyalty. By employing AI-driven segmentation and personalization, companies can meet rising customer expectations for relevant interactions, gaining a competitive edge while fostering stronger, long-term relationships. This project is motivated by the potential to transform customer engagement strategies, enhancing customer satisfaction and ultimately contributing to sustainable business growth.

**1.3 Objectives**

The objectives of this project are as follows:

* To create a recommendation system that can dynamically deliver personalized product recommendations and promotions based on customer segments.
* To design targeted marketing campaigns that address the specific needs of different customer groups, improving engagement and conversion rates.
* To analyze the effectiveness of personalized marketing efforts on key business metrics, such as customer engagement, satisfaction, and purchase frequency.

**1.4 Scope of the Project**

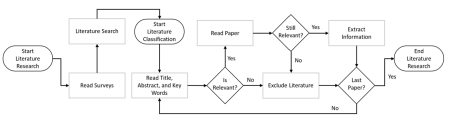
The scope of this project includes:

* Data Collection and Analysis: Gathering customer data, including transaction history, browsing patterns, and demographic information, to serve as input for segmentation.
* AI and Machine Learning Algorithms: Using clustering, to develop AI-based customer segmentation by analyzing customer data such as preferences, behaviors, and purchase history.
* classification, and recommendation algorithms to identify patterns and create segments.
* Recommendation Engine Development: Building a recommendation engine that can dynamically update and provide personalized content to each customer segment.
* Campaign Personalization: Designing marketing campaigns tailored to each segment, including email marketing, digital ads, and product recommendations.
* Evaluation and Feedback Loop: Continuously measuring the effectiveness of the personalized campaigns and refining the AI models based on customer response and feedback.
* This project emphasizes delivering measurable improvements in customer satisfaction, engagement, and long-term loyalty through AI-driven personalization and segmentation.

**CHAPTER 2**

**Literature Survey**

As already encouraged in the introduction we want to scientifically investigate which processes exist for personalized customer marketing approaches. Especially, to get an overview of commonly used customer segmentation methods in the context of CRM in e-commerce, we have conducted an extensive literature review. Thereby, Brocke et al. (2015) published a recommendation on how to conduct such a search in an effective and highly qualified way. Hence, we followed their recommendation for the most part. Figure 1 illustrates our review process. We started our literature research by reading survey papers to derive an integrated and consolidated understanding of the conceptualization of the subject. Thereafter, we started the literature search. Therefore, we our search scope. Vo Brocke et al. (2015)

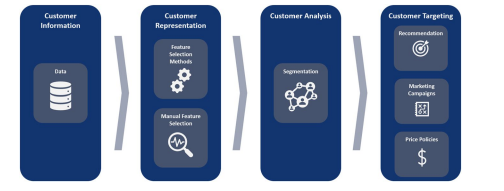


**FIG.1(Flow chart of the literature research process)**

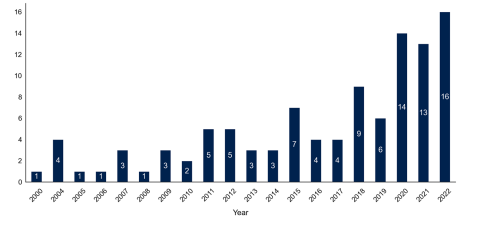
refer to Cooper (1988) which states four steps on how to define a search scope: (1) process, (2) sources, (3) coverage, and (4) technique. Leaning on these four steps we choose a sequential search process. As a publication source, we used the Web of Science1 online research tool as it is one of the leading and analytical platforms and provides scientific publications across a wide amount of knowledge domains (Li et al. 2018). To keep the focus on the customer segmentation methods we used the following search term: • “Customer segmentation” or “customer clustering” or “user segmentation” or “user clustering” Herein, we chose to use the word “user” as a synonym for “customer” and “clustering” for “segmentation”. We wanted the search to be as less restrictive as possible to not miss relevant publications. Therefore, we expected works that are not relevant to our research. After having a corpus of hundreds of publications, we started reading the title, abstract, and keywords of the publications. We out all publications that did not deal withcustomer behavior in commerce, especially in the context of e-commerce. The next step was to read all remaining papers and excluded all publications that did not deal with customer segmentation in an e-commerce use case and it became apparent that customers were segmented based on their information and actions. We extracted all wanted information from publications we classified as relevant., we retrieved bibliometric information, information about the use case, the used methods, information about the used data, and the results.

**2.1 Literature overview**

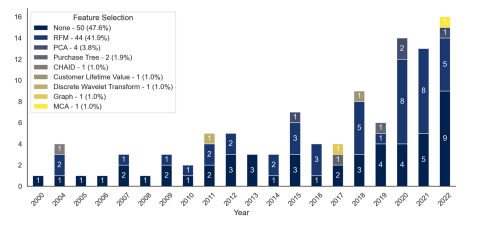
As aforementioned in Sect.  2, we started our literature review with reading related surveys. Plenty of research surveys in the of segmentation prioritize the underlying methodology or class of methods but not their usage in specific domain (Gennari 1989; Rokach 2010; Hirola 2013; Ben Ayed et al. 2014; Firdaus and Uddin 2015; Reddy and Vinz muri 2018; Shi and Pun-Cheng 2019). For example, Shi and Pun-Cheng (2019) review clustering methods for spatiotemporal data which are collected in diverse domains like social media, human mobility, or transportation analysis. Another survey example is brought by Hirola (2013). The author reviews segmentation approaches for applications of soft computing techniques. Other surveys or studies focus on specific methods like k-means or RFM-analysis (Sarvari et al. 2016; Deng and Gao 2020). The most related literature review wefound in our literature search isfrom Sari et al. (2016) which reviews customer and marketing segmentation methods and the necessary data. They identify different segmentation approaches and e-commerce process which coincides in some parts with our outcomes. However, as already mentioned before, six years have already passed and their paper corpus consist of less than 20 publications. From this, we deduce the need for an up-to-date and more detailed review in the area of customer segmentation in e-commerce. The Woos search from 2023/01/01 led to 852 publications, of which not all were related to our research as assumed. As described, we excluded all publication that did not deal with customer behavior in e-commerce.



**FIG..2 (Process of customer targeting based on behavioral information gathered from data)**



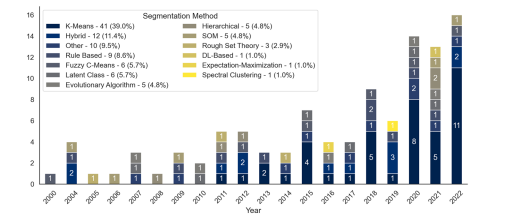
**FIG.3(Distribution of surveyed publications from 2000 until 2022)**



**FIG.4 (Distribution of the surveyed feature selection methods over the year)**

**2.2 Indepth customer segmentation methods**

The authors of the reviewed publications utilize different customer segmentation methods for the customer targeting process. Figure 5 shows the distribution of segmentation methods among all publications and over the years. K-means is the most frequently used customer segmentation method in our surveyed literature (41 of 105). The goal of the k-means algorithm is to partition a set of data points into k segments which minimize the distance between the data. Usually, the Euclidean distance is used.

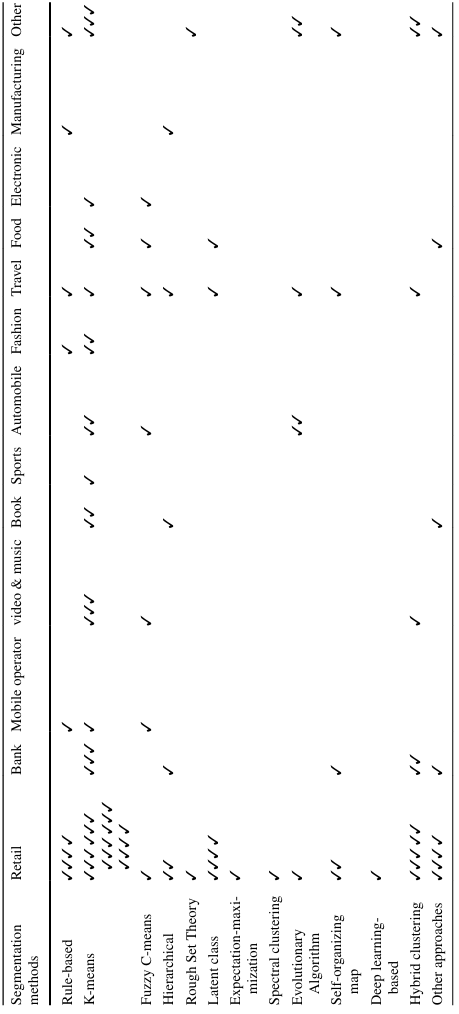


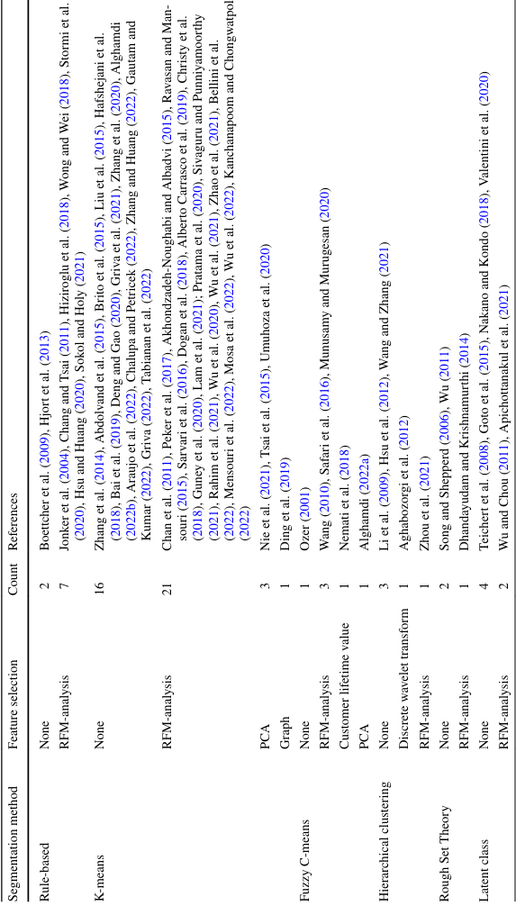
**FIG.5 (Distribution of the surveyed clustering methods over the years of publication)**

**2.3 Overview customer targeting use cases**

The underlying customer targeting process applies to a large amount of business and e-business use cases. In this section, we present an overview of which segmentation methods are used on which use case. Therefore, we brief introduce the found e-commerce use cases. The first category of use cases we want to introduce is Retailing. It is the sale of different goods that are not further specified and don’t belong to any other use case category. We also assign use cases to this category if it is not further specified. This means that a pure sports retailer is classified under the Sports use case, or a retailer that sells only clothing is classified under Fashion. Different to retailing, fashion is a dynamic industry (Brito et al. 2015). Like the fashion branch, Electronic is considered as a branch of e-commerce retailing. In the literature, some customer behavior segmentation use cases are related to Banking. Use cases in this category naturally have more information about the customer. In addition, the products and services don’t change as quickly as in retailing. In Mobile operators’ use cases the authors deal with data from mobile network providers. With YouTube, Netflix, and other companies, Video & music streaming platforms and services become very popular, and forecasts show that sales will also grow strongly in the coming years (Statista. com 2022). In our literature search we found some Book use cases that deal with book retailing or renting services. Nowadays, there a plenty of online services to plan a trip.

**TABLE.1 (Use cases and utilized segmentation approaches for customer targeting of commerce business from the survey literature)**





**TABLE.2 (Segmentation and used feature selection methods with corresponding references of the survey literature)**

**CHAPTER 3**

**Proposed Methodology**

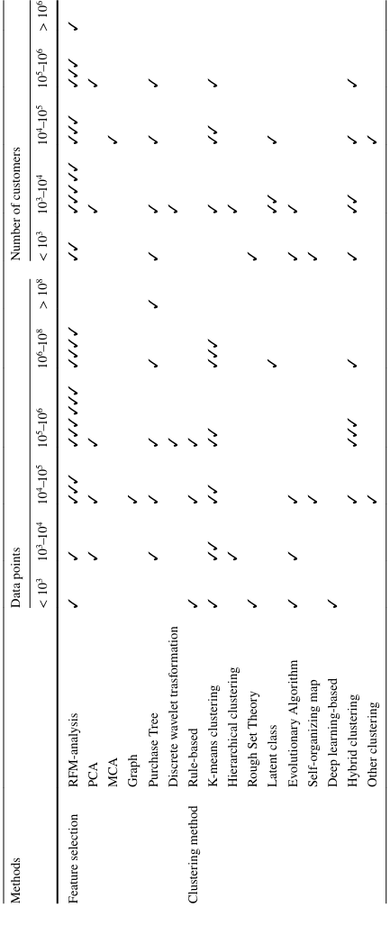
* 1. **System Design**

An upward trend in the number of publications can be recognized which indicates the importance of customer behavior analysis and therefore, their segmentation even after twenty decades of research. Especially, in the years 2020, 2021, and 2022, we have found more publications than the years before. There may be several reasons for this. The first reason that comes to mind is the current covid pandemic. This has increased the growth in e-commerce services. This could have prompted less digitalized companies to digitalize more and offer their services online. In many publications the company remains unknown. However, in some other publications the companies are named

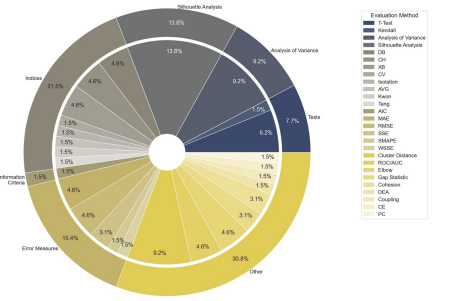
* 1. **Overview and examples of the interplay between customer representation and analysis for customer targeting use cases:**

The authors of the identified publications utilize different customer segmentation methods with different featureselection methods for the customer targeting process. In this section, we further investigate and describe these approaches to give a better insight into the interaction of the feature selection and segmentation methods. Table 2 provides an overview of the different segmentation methods with the corresponding feature selection approaches used. It also lists the number of times such a pair of segmentation method and feature selection was used in the paper corpus. The last column of the table shows the publication’s reference. In the following, we present some examples on how the different segmentation and feature selection methods are used in the found literature to approach customer targeting in e-commerce. In nine publications rule-based clustering is used to segment the customers into different behavioral groups. Therefrom, seven use the RFM-analysis to represent their customers. An example retail use case that combines RFM-analysis and k-means is provided by Hsu and Huang (2020). In their research they want to identify VIP customers. VIP customers are buyers of critical products which are not purchased by the average customer. In their approach, they apply the RFM-analysis on over 600,000 transactions from around 3800 customers.

**.**



**TABLE.3 (Data dimensionality that the feature selection methods and clustering approaches handle in the experiments of the surveyed literature)**



**FIG.6 Distribution of the segmentation methods used evaluation methods.**

**CHAPTER 4**

**Implementation and Result**

**4.1 Implementation**

**Step 1: Data Collection**

**To effectively segment and personalize, businesses must first collect data about their customers. This can include:**

* **Demographic data:** Age, gender, income, education level, etc.
* **Psychographic data:** Values, interests, lifestyle choices, and opinions.
* **Behavioral data:** Purchase history, browsing habits, frequency of interaction with the brand, etc.
* **Transactional data:** Purchase frequency, amount spent, time of purchase, product categories, etc.

**Step 2: Data Analysis**

**Once the data is collected, advanced analytics tools and techniques are used to segment customers into distinct groups. These methods include:**

* **Cluster Analysis:** Grouping customers with similar behaviors or characteristics into clusters.
* **RFM (Recency, Frequency, Monetary) Analysis:** Segmenting based on how recently customers made a purchase, how often, and how much they spent.
* **Predictive Analytics:** Using past customer behavior to predict future behavior and create segments around these predictions.

**Step 3: Customer Segmentation**

**Segmentation can take many forms, but common segmentation approaches include:**

* **Demographic Segmentation:** Grouping customers based on characteristics like age, gender, income, etc.
* **Geographic Segmentation:** Dividing customers based on location or regional preferences.
* **Behavioral Segmentation:** Segmenting customers based on their purchase behavior, online activity, and interactions.
* **Psychographic Segmentation:** Creating segments based on lifestyle, values, interests, and attitudes.

**Step 4: Personalization Strategy**

**Once the segments are created, the next step is developing personalized strategies. Personalization can occur in various forms:**

* **Email Marketing:** Personalizing email content based on customer behavior (e.g., recommending products based on previous purchases).
* **Product Recommendations:** Tailoring product suggestions on e-commerce websites or apps based on browsing history and past purchases.
* **Dynamic Pricing:** Offering customized pricing or discounts based on customer segments, loyalty, or previous buying behavior.
* **Content Personalization:** Customizing website content, including banners, promotions, and product displays, to align with the interests and behavior of each segment.

**Step 5: Automation and Execution**

**Automation tools are essential to implement segmentation and personalization at scale. This can include:**

* **Marketing Automation Software:** Tools like HubSpot, Salesforce, or Marketo to trigger personalized email campaigns based on customer activity or segment.
* **Customer Relationship Management (CRM):** A CRM system to track customer behavior and manage segmentation efforts.
* **AI and Machine Learning:** These technologies can automate personalization by analyzing vast amounts of customer data and adapting offers and content in real time.

**4.2 Results of Customer Segmentation and Personalization**

**Implementing segmentation and personalization has yielded significant benefits for businesses across industries. Some of the most notable results include:**

**Increased Customer Engagement**

* **Higher Open Rates:** Personalized email campaigns generally have higher open and click-through rates than generic ones. Customers are more likely to engage with content that is relevant to them.
* **Improved Website Interaction:** Personalized content on websites increases user engagement and time spent on site.

**Enhanced Customer Satisfaction and Loyalty**

* **Tailored Experiences:** By offering products, recommendations, and content that align with individual preferences, businesses can improve customer satisfaction and retention.
* **Loyalty Programs:** Personalized loyalty offers, such as discounts and rewards based on customer behavior, increase repeat purchases and brand loyalty.

**Increased Conversion Rates**

* **Targeted Offers:** Personalization can drive more targeted sales and improve conversion rates, as customers are more likely to buy products tailored to their needs and preferences.
* **Cross-Selling and Up-Selling:** Personalization increases the likelihood of successful cross-selling and up-selling, as customers are presented with relevant complementary products or premium options.

**Higher Revenue**

* **Product Recommendation Engines:** Businesses have seen a boost in sales by implementing recommendation engines that suggest products customers are likely to purchase based on their previous interactions.
* **Dynamic Pricing:** Offering personalized discounts or pricing based on customer segmentation can drive more sales and improve profit margins.

**Data-Driven Insights**

* **Better Marketing ROI:** By focusing marketing efforts on specific customer segments, businesses can allocate resources more efficiently, resulting in better ROI.
* **Increased Customer Lifetime Value (CLV):** Personalization and segmentation increase the value of each customer over their lifetime by ensuring they continue to find relevant products and experiences.

**CHAPTER 5**

**5.1 Analysis and discussion**

* As previously shown in Fig. 3, the reviewed publications were not equally distributed over the years. An upward trend in the number of publications can be recognized which indicates the importance of customer behavior analysis and therefore, their segmentation even after twenty decades of research. Especially, in the years 2020, 2021, and 2022, we have found more publications than the years before.
* There may be several reasons for this. The first reason that comes to mind is the current covid pandemic. This has increased the growth in e-commerce services.
* This could have prompted less digitalized companies to digitalize more and offer their services online. In many publications the company remains unknown. However, in some other publications the companies are named. Two examples are taobao.com or nelly.com that are established online companies which is an indication against our statement. literature conducted experiments did not show the state of digitization of the companies.
* Therefore, whether this connection exists remains open, and is not further investigated by us. Another reason, and in our opinion a more decisive one, is the increasing availability of the internet regardless of location. This means that a user can access the available online services at any time and from any place. For example, watching a series during a train ride or buying a new product at the online retailer of choice. With new requirements and necessities, the topic is also becoming more relevant in science and thus more is being published.

**5.2 Analysis of feature selection methods**

Based on our research, feature selection to represent customers is a fundamental step in the customer targeting process. For feature selection, customer information is indispensable. It is a challenge to get customers’ demographic information, physiographic information, or information about their preferences. As already stated, there are two possible ways to collect such data. Explicit information collection is done by questionnaires or user surveys that require customers’ accommodation to participate. Another, more implicit way is to collect demographic information via registration. Information can be collected by setting them as mandatory.

Never the less, collecting data via registration is often limited to the usual information like age, gender, or address. In some use cases, like fashion, additional information about height and weight can be collected. It needs to be considered, that some users don’t want to provide any information and wish to remain anonymous. They either give false information or leave the website (service). In both cases, it is not possible to gather useful information and in the worst case, the former leads to false conclusions regarding the customers.

Furthermore, user groups that don’t participate in a survey or are signed up are not represented in the data which makes the acquisition of unknown and new customers harder. It is possible to gather customers’ preferences with the aforementioned method. Nevertheless, this comes with a huge disadvantage. The information is outdated soon and needs to be constantly updated which increases the maintaining effort. Constantly asking the customer for an information update can also cause him to quit as a consequence. Therefore, customer preference should be estimated based on their recent behavior. Customer behavior information can be recorded implicitly. Usually, purchase information with product information, timestamp, etc., is stored for a company’s financial overview. In addition, online touchpoints with the customer can be logged by the system. These logs can include various touchpoints like product views, click **shopping** events, reviews, (dis)like, and many more. The advantages are that the customers do not disclose any personal information. Also, they are likely not interrupted on their journey by unwanted questions. Nevertheless, disadvantages exist too. Predicting customer information from their behavior is not always correct that is for example caused by customers’ heterogeneity.

Additionally, a large amount of data is required to make such predictions. Another challenge of implicit data collection is that the information needs to be linked to the customer. However, there are plenty of tracking-techniques to link the data with customers by using cookies or the browser identifier to name two examples. As shown in Fig. 4, for the customer process as a whole, it makes no difference whether a feature selection method is used or the features are selected or handcrafted by an expert.

However, manual feature selection and feature selection methods have their pros and cons. One advantage of manual feature selection is that no additional computation is required. However, it requires expertise and domain knowledge to select customer information that is meaningful and representative. Feature selection methods are designed to automate the selection of features. One advantage is that domain knowledge is no longer required. However, this doesn’t mean that domain knowledge should generally be dispensed with. Another argument in favor of feature selection methods is that information redundancy can be removed.

Redundancies come in hand with the amount of data collected. Removing unnecessary and redundant information can speed up the customer analysis algorithms. This information is hard to determine and select manually even with domain knowledge. Regarding Table 3, we notice that feature selection methods have processed larger amounts of data in our literature. Considering our second question from the introduction, we can state that feature selection methods allow larger amounts of data forcustomer behavior analysis. Particularly, the RFM-analysis and Purchase Tree have no limitation concerning the data dimensionality based on our research.

Our literature research shows that the RFM-analysis is by far the most popular feature selection method. Therefore, we analyze the RFM-analysis method in more detail hereafter and discuss the advantages and disadvantages. During the literature research, several points caught our attention. The RFM-analysis could be applied to almost any type of purchase or activity data since only three features need to be calculated.

Furthermore, the calculation is very simple and requires only the basic arithmetic operations, so there is valuable customer representation in only three values. These values can be represented either numerically or categorically. For the categorical representation, the values were typically divided into five categories, each with 20%-quantiles. Thus, the obtained features are used for any clustering method. In addition, we notice that the RFM-analysis is often extended with additional features.

The feature extension is usually use case-specific. Besides adding new features, the RFM-features are extended on different activity levels. For example, the RFM-values are calculated for all product categories or different customer activities. This provides additional information about the customer’s product preference at the category or activity level. Another advantage of RFM-analysis is that it can handle all sizes of data sets without having a scalability problem. This has been sufficiently demonstrated in the publications and is illustrated by Table 3. We also like to note that in some publications, the RFM-analysis is used to explain the resulting clusters and helps with the customer behavior analysis which shows that decision makers can easily understand and interpret the RFM-values. Based onour findings to feature selection methods, we can answer the third question as follows. For feature selection methods no time-depended methodological trend could be determined. However, the most popular feature selection method is the RFM-analysis.

**5.3 Analysis of segmentation methods**

We found 13 different types of segmentation methods. K-means is by far the most used approach. Especially, in the last years from 2020 to 2022 k-means is used 24 times. In regard to the third guiding question, we can conclude that besides a k-means upwards trend no other trend can be spotted. The question that now arises is “why is k-means becoming so popular recently”? One answer is that k-means is simple to implement and an established approach. In contrast, other approaches like EAs, hierarchical clustering, or SOMs are more complex according to how the run time or space requirements grow as the input size grows (Bachmann-Landau notation) and it needs more effort to implement them (Firdaus and Uddin 2015).

The ever-increasing amounts of data in e-commerce amplifies this trend because simple methods can be used more quickly, and thus, results can be obtained faster. However, if this is the reason, then the question that follows is why are rule-based approaches not popular as well? As shown by Fig.  5 the density of rule-base approaches increased in the years between 2018 and 2021 but some other influencing factors play a major role on the methods popularity. While we can only make assumptions at this point, rule-based segmentation approaches have significant drawbacks. For example, they require domain knowledge to set appropriate thresholds for separating customer segments.

The increasing and heterogeneous amount of data complicates this setting of appropriate thresholds or requires a higher dynamic, which in turn results in more rules and complex relationships.

Considering the data dimensionality which is used in the publications we see that k-means approaches can handle a larger amount of data and is in pair with latent class approaches. As we **mentioned, the** hybrid approach that uses the largest amount of data is a combination of the latent class model. However, concerning the number of customers in the data which are the objective of the clustering, the numbers rarely exceed the 10,000. This indicates that clustering approaches need an appropriate feature selection method to deal with a larger amount of data. All this doesn’t mean that the methods cannot be applied to larger data sets. Our argumentation is based solely on the paper corpus we saw.

Based on the findings concerning the data dimensionality, we can state for guiding question number three, that k-means and latent class models can process the largest amount of data among all segmentation methods. However, as already stated this applies only in case of manual feature selection. We recommend using a namely the RFM-analysis that allows to process any kind of data dimensionality. Note that we don’t address the time or memory complexity of the segmentation methods, which is also a performance indicator, but evaluate them based solely on the amount of data used in the literature. In terms of use cases, we can state that each clustering method is usable in retailing use cases.

We cannot make such a generalized statement for other domains. However, it is not unlikely that all segmentation methods can be used independently of the domain. Especially with k-means, we can see that it has the largest variant of different use cases. Nonetheless, the reason for being used in different domains can be because k-means is applied in most publications.

Apart from a quantitative analysis of the segmentation method, we would like to make a qualitative analysis. Unfortunately, there is no way to determine which segmentation method performs best. The major issue in our opinion is that there is no ground truth for the customer segments to determine a score. Therefore, there is no unified method for qualitative evaluation which is necessary to state which segmentation method is superior to the other. We noticed that there a vast amount of different evaluation methods as presented in Sect. 3.6. Different evaluation approaches are required for different clustering approaches, i.e. fuzzy (soft) clustering has different properties than hard clustering. It would simplify qualitative segmentation analysis if the scientific community agree on a small set of evaluation methods.

The urge is there which we can see in the number of different evaluation metrics and the considered publication where the authors try to show that their approach is superior to others. If everyone would use the same metrics, the authors’ efforts would have more significance and the performance of the method could be compared over different publications which are usually done in other scientific disciplines. Nevertheless, due to the absence of ground truth, correctness can never be shown, and therefore, the purpose of unified evaluation methods may be questioned. Another aspect we want to consider is evaluation metrics with semantic interpretability. Such metrics would have the advantage to show which segmentation algorithm partitions the customers in a desirable way. Furthermore, it would create comparability between multiple segmentation methods for identical use cases. However, the challenge is to define evaluation metrics that have the capacity to be semantic interpretable and, at the same time, can be applied to different segmentation methods and use cases.

**5.4 Challenges and Best Practices**

**Challenges**

* **Data Privacy Concerns:** With increasing concerns about privacy, businesses need to ensure they comply with data protection laws like GDPR and CCPA while gathering and using customer data.
* **Data Integration:** Businesses often face challenges in integrating data from multiple touchpoints (e.g., websites, apps, physical stores).
* **Over-Personalization:** There is a risk of overwhelming customers with too many personalized messages, which can lead to frustration or fatigue.

**Best Practices**

* **Use AI and Machine Learning:** Leverage AI to analyze customer behavior at scale and deliver dynamic personalization in real-time.
* **Maintain Transparency:** Ensure customers understand how their data is being used and offer easy opt-in/opt-out options for data collection.
* **Test and Optimize:** Continuously test personalization strategies through A/B testing to refine offers, recommendations, and messages.

**6 Conclusion and future research**

In this survey, we provided an extensive literature review on customer targeting process for e-commerce use cases whose main focus lies in the segmentation methods for customer behavior analysis. Our goal was to provide an overview of segmentation methods used in the literature and to determine best-practice approaches and their limitations. We introduced the steps of the research and key criteria for the paper selection and analyzed as well as our findings afterward. In our work, we considered 105 publications with different case studies that focused on customer analysis with segmentation methods. Summarizing the approaches examined, the identified four-step process emerges as the current gold standard for personalized customer targeting in e-commerce. For the customer representation, either hand-crafted features or an RFM analysis adapted to the use case are generally used. Subsequently, for customer analysis, the generated customer representation is segmented using a k-means approach. Based on our research and literature analysis we made several findings regarding our investigated topic.

• We identified a common process for personalized customer targeting which includes feature selection methods, customer segmentation, and customer targeting. This process is illustrated by Fig. 2 and can be utilized to plan customer targeting campaigns. Each of the four steps has its own requirements and it a discipline of its own worth to be investigated. We focused on the customer analysis and customer representation part.

• Over the years, the number of publications that deals with customer targeting in e-commerce are continuously increasing. This supports the preceding assumption that it is a time-relevant subject.

• Feature selection methods enable the usage of larger datasets and among the utilized methods the RFM-analysis is by far the most popular one. There are many reasons for this: first, the method is easy to use, and second, it is based on features that can be extracted and understood. Another advantage of RFM analysis is the possibility of its easy adaptation to specific use cases by adding further or changing existing features.

• In approximately half of the publications (47.6%), manual feature selection was used.

• Among all the used clustering methods, k-means has emerged as the most popular approach (39% in total). Since 2011, it was repeatably used. Besides that, no other

canover-time trend identified. The popularity of k-means can be explained by its simplicity and applicability to large scale datasets.

• We were not able to define the best clustering approach based on its performance because many different evaluation methods exist and were used to evaluate the cluster quality.

• Some evaluation methods can be used to determine the optimal number of segments which is unknown from the beginning and is often a tunable hyperparameter.

• The literature review doesn’t show that a segmentation method exists that is applicable to every e-commerce use case that involves customer analysis. This could only be suggested, if at all, for the retail use case. In terms of method, k-means has been used in every use case identified, with the exception of the manufacturing use case. New insights always come with new challenges and opportunities. Based on our research and findings we propose future research ideas which should be investigated. Especially with regard to recent developments in the Feld of Deep Learning, there are many approaches that can be adapted and, according to the assessment, display a lot of potential.

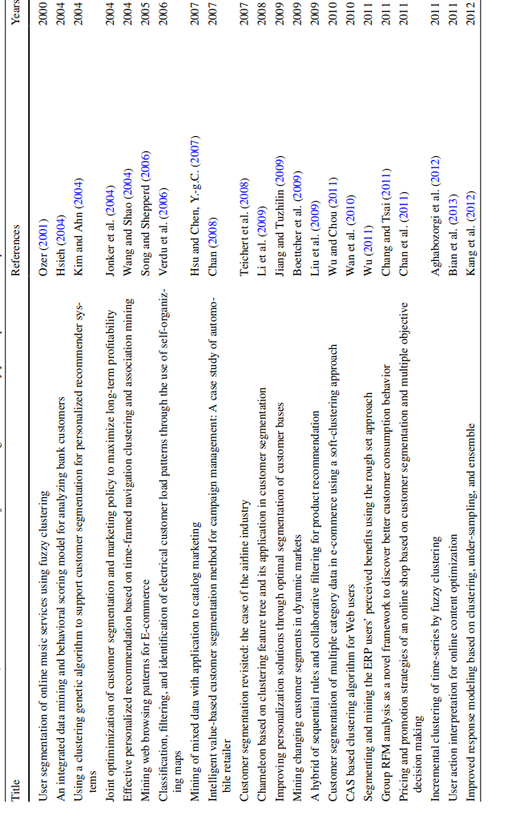
• Deep learning introduced innovations in many domains such as natural language processing and computer vision. Nevertheless, we only found one DL-based segmentation approach in our research. Therefore, we see potential and a research gap in DL techniques for segmentation.

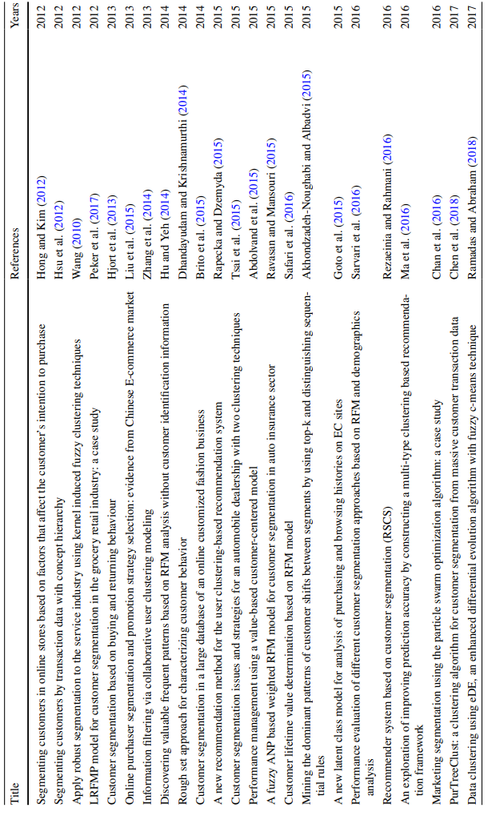
• The process steps in the identified four-phase process for customer targeting are essentially based on a high level of understanding of the customers, i.e. their needs and behavior. This is necessary for marketing and domain expert to tailor personalized marketing strategies for the customers. However, with the advent of deep learning-based approaches personalized customer targeting can be done fully automated e.g. end-to-end model and therefore, the customer analysis step which includes customer segmentation be omitted. This development can be seen for example in deep learning-based recommendation systems which make personalized recommendation without the need of the customer analysis. This leads to the question; How customizable are the individual phases of this process and can individual steps be omitted to increase efficiency or are all steps so fundamental that a deviation from these procedures would have a negative impact on the goal, customer targeting?

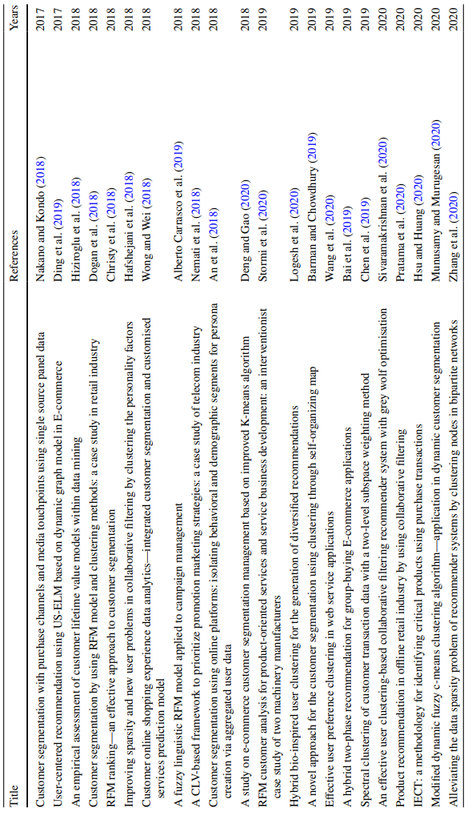
• Manual feature selection is still frequently used. The feature quality is thereby highly depended on the underlying expertise to select or define important features for clustering. Progressive digitization is leading to growing challenges, especially in dealing with data volumes and data diversity. To meet these challenges, manual feature selection is reaching its limits as it is not able to tap the insight potential within this data. Hence, the question arises if approaches exist that can help experts to create meaningful and representative features for customer representation? In this regard a look outside the box to other e-commerce research, e.g. clickthrough rates prediction can yield new approaches. There researchers and professionals have started using feature embeddings on manual selected features with the underlying assumption that the learning models will learn meaningful representations from the data. This would simplify the manual feature selection process. However, these learning models are usually based on deep neural networks which are unfortunately black boxes and not interpretable. The question rises, if segmentation methods can be used as a post-processing to provide interpretability for the embedded features and therefore, an insight over the customers? (Which got lost by not using the customer analysis step).

• In our research, we identified many different evaluation metrics to evaluate the performance of segmentation methods. Nevertheless**,** we could not find a consensus on evaluation metrics as in other domains. The reason is the missing ground-truth. This circumstance makes it difficult to determine the effectiveness andtransferability of a segmentation approach from one use case to another. The open question that remains is, is it necessary, to develop evaluation metrics with semantic meaning and is it possible to transfer such metrics to different experiments to enable comparison of the segmentation methods? In our literature review, we covered the usage of feature selection and segmentation method for personalized customer targeting. E-commerce is a dynamic environment with ever new challenges and therefore, new research opportunities.

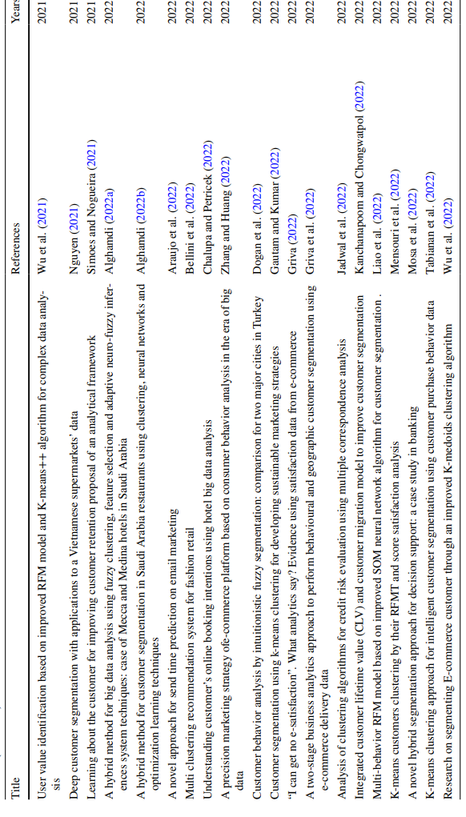
**TABLE.1 (Literature with title, author and date which are the object of investigation sorted by year of publication/acceptance)**











**REFERENCES**

* Zhang S, Yao L, Sun A, Tay Y (2019) Deep learning based recommender system: a survey and new perspectives. ACM Computed Surb (CSUR) 52(1):1–38 Zhang F, Qi S, Liu Q, Mao M, Zeng A (2020) Alleviating the data sparsity problem of recommender systems by clustering no
* Wang Y, Zhou J-T, Li X, Song X (2020) Effective user preference clustering in web service applications. Computed J 63(11):1633–1643. <https://doi.org/10.1093/comjnl/bxz090>
* Tsai C-F, Hu Y-H, Lu Y-H (2015) Customer segmentation issues and strategies for an automobile dealership with two clustering techniques. Expert Syst 32(1):65–76. https://doi.org/10.1111/ exsy.12056
* Stormi K, Lindholm A, Laine T, Korhonen T (2020) RFM customer analysis for product-oriented services and service business development: an interventionist case study of two machinery manufacturers J Manag Gov 24(3):623–653. <https://doi.org/10.1007/s10997>
* Sari JN, Nugroho LE, Dedriana R, Santosa PI (2016) Review on customer segmentation technique on ecommerce. Adv Sci Lett 22(10):3018–3022
* Davies DL, Bouldin DW (1979) A cluster separation measure. IEEE Trans Pattern Anal Mach Intel PAMI 1(2):224–227. <https://doi.org/10.1109/TPAMI.1979.4766909>
* Gautam N, Kumar N (2022) Customer segmentation using k-means clustering for developing sustainable marketing strategies. Bizen’s Inf-Bus Inf 16(1): 72–82. [https://doi.org/10.17323/2587-814X.2022.1. 72.82](https://doi.org/10.17323/2587-814X.2022.1.%2072.82)
* Krishna GJ, Ravi V (2021) High utility itemset mining using binary differential evolution: an application to customer segmentation. Expert Syst Appl. <https://doi.org/10.1016/j.eswa.2021.115122>
* Ma X, Lu H, Gan Z, Zhao Q (2016) An exploration of improving prediction accuracy by constructing a multi-type clustering based recommendation framework. Neurocomputing 191:388–397. https:// doi.org/10.1016/j.neucom.2016.01.040
* Nakano S, Kondo FN (2018) Customer segmentation with purchase channels and media touchpoints using single source panel data. J Retail Consume Serv 41:142–152. [https://doi.org/10.1016/j.jretc onser.2017.11.012](https://doi.org/10.1016/j.jretc%20onser.2017.11.012)
* Rousseau PJ (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Computed Appl Math 20:53–65. <https://doi.org/10.1016/0377-0427(87)90125-7>