ISSN 1870-4069

# Towards an Intelligent Shop Keeper-Centric Transformation

Juan Arturo Pérez-Cebreros[1,2], Gabriel Sánchez-Pérez[1],
Luis Mario Hernández-Rojas[1], Ramón Mendoza-Hernández[1],
Emiliano Lorences-Gutiérrez[1]

[1] Instituto Politécnico Nacional,
ESIME Culhuacán,
Departamento de Computación,
México

[2] Universidad Jesuita de Guadalajara,
Departamento de Electrónica,
México

juan.perezc@iteso.mx, gasanchezp@ipn.mx,
{lhernandezr1800, rmendozah1600,
elorencesg1800}@alumno.ipn.mx

**Abstract.** Intelligent systems are transforming, in the sense that they are no longer novelties used by some companies, but are now serious business tools that shape the digital world. During the pandemic, different companies sought to digitalize a traditional approach by means of an ecosystem of technological and logistic solutions, whose objective is to strengthen shopkeepers and make them more competitive, applying different models for supply purposes. In this sense, our study is important because it can help grocery wholesalers sustain shopkeepers through face-to-face and digital supply solutions. Adequate categorization of the shopkeeper group is necessary to achieve successful development of marketing strategies that will align perfectly with each group. In this research, we select a Content Management System (CMS) and incorporate a clustering extension with the aim of performing an exploratory analysis of the behavior of different variants pertaining to the K-means algorithm (Hartigan-Wong, Lloyd, Forgy and MacQueen) in the context of a data set of shopkeepers from Mexico City. By means of these Artificial Intelligence techniques, we focus on the prediction of cross-selling opportunities and try to answer questions about who are the most likely shopkeepers to buy additional products, in order to propose an intelligent transformation model focused on the corner shop and study this new way of digitally supplying street stores.

**Keywords:** Cluster analysis, content management system, shopkeepers.

## 1 Introduction

During this global pandemic known as COVID-19, shopkeepers or small corner stores, consisting of those small businesses better known as convenience or grocery shops that are present in residential areas and provide products from the basic food basket, sustain more than 3 million families, are established throughout the country and make large contributions to the national economy.

However, with the arrival of the pandemic, new challenges arose for these small businesses that have put their economic stability at risk.

The importance of digitizing shopkeepers in times of pandemic is demonstrated by emphasizing that these types of stores must take advantage of and update themselves by employing the various digital tools available in order to reach consumers more effectively, as not only can they compete more effectively, but can more easily identify their market, thus precluding that this type of business would close due to the pandemic.

In [1], a platform called "*mandamelo*" is being created for the benefit of corner stores and consumers, making it possible to sell online. In [2] opines that it is necessary for businesses to adapt quickly in order to immediately apprehend the needs of the consumer. For this reason, different ventures have emerged, such as "*Rabbit*" and "*Nutenta*" that offer an application so that you can manage your orders digitally, thus making procedures more efficient, by easing communication and decentralization.

This increase in shopkeeper-centric platforms leads to greater competition, making it necessary to study their different personalities, geographical positions and preferences.

This means that having the same strategy for all shopkeepers is not sufficient [3], so it is vital to be able to categorize them and distinguish the differences between each type of shopkeeper. The aim of this research is to expand Magento in order to focus on finding the most successful groups of shopkeepers among shopkeepers as a whole, while also improving purchase suggestions.

## 2 Literature Review

The classification and identification of patterns from customer data is very important to support business decision-making [4]. In the marketing context, grouping methods become powerful tools that make it possible to categorize customers in order to identify their patterns and behaviors, in terms of the products or services offered by companies [5].

In [6], the two most popular partition-based clustering algorithms, K-Means and K-Medoids, are evaluated using the transactional dataset. Results from the comparison show that the time spent in the selection of the initial values and the spatial complexity of the cluster overlap is much better in the case of K-Medoids than in the case of K- Means.

Furthermore, K-Medoids is superior in terms of execution time, not sensitive to outliers, and reduces noise compared to KMeans, as it minimizes the sum of the differences of the data objects.

In their research paper, Doğan et al. [7] performed a customer categorization for one of the largest sports retail chains in Turkey, based on the RFM model. The purpose of this research was to find new clusters that would help to redefine the existing card loyalty system, for which they used a data set made up of the Recency, Frequency and Monetary (RFM) indicators for 700032 customers, who made purchases either in person or online.

Two methodologies were applied; firstly, two-stage clustering, from which the Bronze client groups were obtained, with RFM indicators below average; Gold, with an R indicator above average and F and M indicators below, and Premium, with RFM above average; and secondly, clustering by k-means, from which groups of Regular clients emerged, which included 92% of clients and with RFM indicators below average; Loyal, with above-average RFM indicators; Star, with RFM indicators well above average and representing less than 0.015% of customers; and Advanced, with RFM indicators above average but less than those pertaining to the Loyal group.

Tavakoli et al. [8] produced a research article, where the idea was to categorize customers using the K-Means method based on an R + FM model that compared this to traditional RFM, taking into account business changes that make it more effective.

The procedure was applied to Digikala, the largest e-commerce in the Middle East, for which four categories were obtained: assets with high value, medium assets with high monetary value, medium assets with high frequency and low activity value, for which they built and applied marketing strategies for each category.

Results from the various initiatives showed that the new categorization model generated greater impact on customers and therefore had greater effect.

In [9] the importance of cluster analysis in the retail industry is mentioned; in order to identify customers according to their purchasing habits, patterns and behaviors, and thus improve customer service and customer satisfaction, resulting in increased allegiance. K-means clustering results help create a business intelligence system, by providing powerful multidimensional analysis and visualization tools, including building sophisticated data cubes with reference to data analysis needs.

Using a business intelligence application that incorporates this grouping system, as a mechanism to manage a retail business, will provide retailers with the means to categorize customers and better understand their behavior and needs, while making knowledge-based decisions in order to provide a personalized and efficient customer service.

Wu et al. [10] wrote a paper that explains the results obtained, when performing customer categorization in a company dedicated to electronic commerce in Beijing, China. During the investigation, the company's transactional data was analyzed, and then the k-means algorithm fused with the RFM model was applied.

As a result, four clusters categorized according to their purchasing habits were obtained, for which a different CRM strategy was developed in order to achieve a greater level of customer satisfaction. Finally, the company's KPIs improved, verifying the effectiveness of the method applied, as: the number of active customers increased by 519, the purchase volume increased by 279% and total consumption by 102%.

## 3 Content Management System under Study

The e-Commerce platforms consist of a software system (CMS) that permits you to manage the contents, mainly those of the product to be sold, and through templates provide design to the visual aspect of the online store [11].

Sometimes, depending on the platform chosen, it will not be necessary to have programming knowledge as, thanks to the use of templates, only the specific characteristics required will have to be incorporated.

In this section, we provide details of the selected tools that were studied: Magento, Woocommerce, and Shopify.

a) Shopify is one of the most commonly used ecommerce platforms worldwide, particularly as a result of its versatility when creating a virtual store. Its extensive options for organization and personalization of both products and the store in general, as well as the different processes it offers for payment and the tracking of orders, represent some of its strong points [12].
b) Woocommerce is the WordPress plugin that can turn a website into an eCommerce. Installation is accomplished with a couple of clicks, without requiring knowledge of programming. Once installed, you can add products, create categories or assess shipping costs [13].
c) Magento, in addition to being a recognized CMS or generating web pages, offers its users thousands of alternatives that include methods for payments, shipments, taxes, analytics and logistics; essential for competing in electronic commerce. Basically, it offers a comprehensive open code solution for the development of web pages aimed at online sales; however, for people who lack programming knowledge, Magento is not easy to use [14].

Extending the work in [15], where 36 features were evaluated and compared for each CMS, it was concluded that Magento should receive the best score, followed immediately by Shopify. However, it does not have Artificial Intelligence features that would permit extended cross-selling and up-selling functions. For this reason, in this investigation, we propose an extension to this platform; incorporating automatic learning mechanisms that make an intelligent sale possible.

Importantly, this represents an applied type of research because it intends to solve a real problem and is a non-experimental cross-sectional descriptive design because the data used was limited to a specific period of time and its purpose was to discover customer groups based on their diverse purchasing habits.

The people were all shopkeepers, who made at least one purchase using the e-marketplaces in the study. Likewise, because this project intends to discover hidden patterns based on a considerably large amount of data, 925 customers, who made one or more purchases during the months of May to August 2022 in the e-marketplaces, were selected as a sample.

## 4 Experimental Procedure

To solve this problem, a three-phase model was proposed:

a) Magento extension phase. In the first phase, we proposed making direct use of the Magento tables using an extraction, transformation and loading process that was carried out in Python.

In Figure 1, we can see part of the Magento data model, which represents a fairly standardized model, but we can access this directly and make use of its master tables and catalog tables, in order to employ data related to customer orders.

This is instead of employing the platform's endpoint or web services because doing so would be impractical and slow.


**Fig. 1.** Diagram for Magento Tables.

The reason for this extension is to provide Magento with intelligent mechanisms for cross-selling and customer categorization, which in our case will focus on shopkeepers. Some previous analyzes [16, 17] indicate that the owners of these convenience stores play a very important role, especially in relation to their personalities.

Therefore, the relationship that is generated between suppliers, customers and their employees promotes commercial development.

b) Shopkeeper categorization phase. Extending the work in [18], we consider the market of a group of shopkeepers based on multiple criteria:

1. $P=\{p_1, p_2,\ldots p_n\}$ is the set of n products.
2. $C=\{c_1, c_2,..c_k\}$ is the set of k evaluation criteria.
3. $T=[t_1,t_2,..,t_j\}$ is the set of j shopkeepers who participate in the market.

These evaluations are used and, with the help of an addition function, a ranking for their products is created. However, this solution is unrealistic due to the large number of shopkeepers and products to be evaluated in the market.

Our proposal is to carry out this ranking automatically. For this, we need to know certain data a priori: we need to know all the products (a*, b*) that compete with each other (this competition will be our criterion).

This will enable us to make comparisons in the form of pairs between each one of the products from the same category (competitive criteria).

To do this, we can make use of the records that are stored in the order tables and automatically establish certain preference rankings for each shopkeeper.

**Table 1.** Set of comparisons between pairs of products.

| Categories | Comparisons between pairs of products |
|---|---|
| $CP_1^{t25}$ | p11 > p20 > p14 > p34 > p10 > p20 > p32 > p52 > p42 > p63 |
| $CP_2^{t25}$ | p41 > p52 > p56 > p64 > p36 > p77 > p42 > p62 |
| $CP_3^{t25}$ | p67 > p87 > p97 > p66 > p98 |

To obtain this ranking, we must first define and rank all the comparisons for each product pair (CP) that compete with each other. Let me emphasize that for each $CP(p_i,p_j)_s\ t_r$ where s=1,..,K, in the maximum number of competing subsets (categories). For example, the purchase data for the shopkeeper $t_{25}$ provides us with certain product comparisons ($p_i$, $p_j$) and indicates to us that three categories exist, in which we find competing products for sale.

In table 1, we can see the product rankings that result from the comparisons of each pair of products that belong to the same competition criteria. This competition criteria or category may constitute different brands of soft drinks that compete with each other. Thus, the products in a category will be defined as the subset of all the products that compete with each other and for each shopkeeper $t_r$, we will obtain a subset of possible rankings that represent their preferences.

For example, the set of possible rankings for the shopkeeper $t_{25}$ is made up of three categories and a totality of 23 products that he buys to sell in his store. Therefore, the ranking of each category makes it possible to identify star products and also those least sold for each category.

We propose market categorization based on preferences and direct commercial strategies for each category of shopkeepers. For this purpose, we must first define the degree of similarity in terms of shopkeepers' preferences. This will allow us to establish better trading strategies.

c) Analysis phase for computational experiments. Today, a number of e-commerce companies produce large quantities of data. Most of the time, this data reveals hidden patterns that can be very useful for decision making. One of the ways to acquire new knowledge or find hidden patterns in the data is by applying categorization or clustering algorithms [19]. In [20], shows that the most widely used clustering algorithm is K-means because its results are easy to interpret and there are different implementations.

Something to consider is that in the literature [21, 22, 23], there are data sets where the K-means does not work in the desired way, because the centroids must be adequately separated. For this reason, in our research we carry out a series of experiments with the aim of obtaining a preliminary exploratory analysis of the behavior of different variants among k-means algorithms.

During this phase, we employed R language, which allowed us to implement four variants for this algorithm. These variants are Hartigan-Wong, Lloyd, Forgy, and MacQueen [24, 25]. K-means algorithms include a large number of variants; however, there are few comparative studies.

In this sense, in [25, 26, 27] they implemented variants of these algorithms in Python and R but for synthetic instances in the FCPS repository [28]. In contrast, we select and

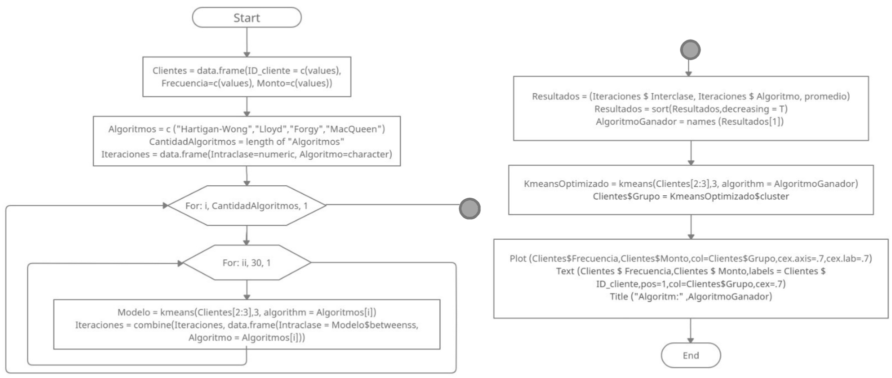


**Fig. 2.** Diagram for each K-means variant.

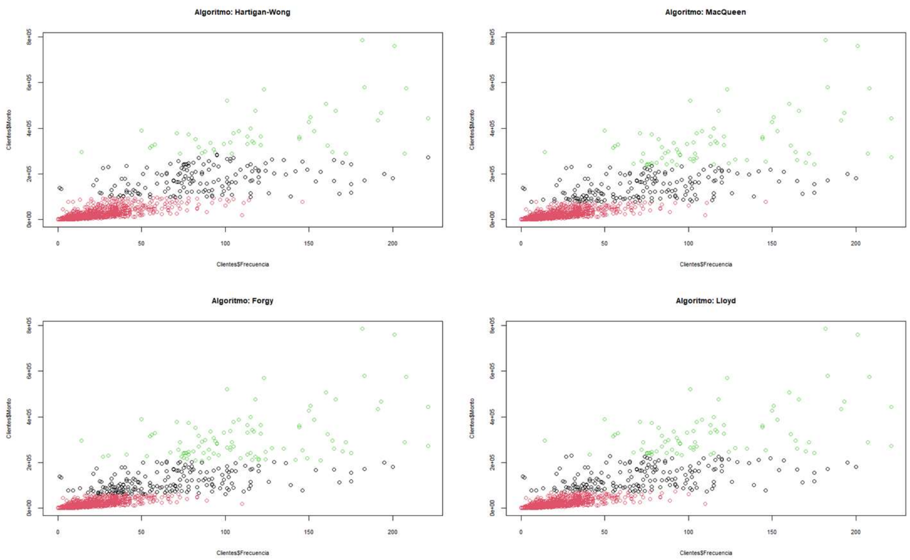


**Fig. 3.** Graphs showing each of the four variants for the K-means algorithm.

compare the four variants provided by the R language and apply them to our dataset of shopkeepers.

This comparison was made using the intra-cluster distance, which equals the sum of the distances between the centroids. The variant of the algorithm that shows the greatest intra-cluster distance will be the one that best separates the clusters.

Above all, the K-means algorithm is an interactive method that implies dividing a set of n objects into k >= 2 clusters, so that the objects in each cluster are similar to each other but different to the objects in the other clusters [29].

More precisely, the problem that the K-means algorithm solves can be formalized as follows: Let N= $\{t_1,..t_n\}$ where the set of n objects will be the set of shopkeepers, who will be divided and categorized by applying a similarity criterion, where $x_i \in R^d$ for

**Table 2**. Results from the thirty implementations.

| Hartigan-Wong | Lloyd | Forgy | MacQueen |
|---|---|---|---|
| [ 1 ] 9.385968e+12 | [ 1 ] 9.281242e+12 | [ 1 ] 9.27003e+12 | [ 1 ] 9.344189e+12 |
| [ 2 ] 9.385968e+12 | [ 2 ] 9.259906e+12 | [ 2 ] 9.201645e+12 | [ 2 ] 9.294131e+12 |
| [ 3 ] 9.385968e+12 | [ 3 ] 9.169836e+12 | [ 3 ] 9.169836e+12 | [ 3 ] 9.281285e+12 |
| [ 4 ] 9.385968e+12 | [ 4 ] 9.273601e+12 | [ 4 ] 9.241439e+12 | [ 4 ] 9.272078e+12 |
| [ 5 ] 9.385968e+12 | [ 5 ] 9.263678e+12 | [ 5 ] 9.176478e+12 | [ 5 ] 9.288699e+12 |
| [ 6 ] 9.385968e+12 | [ 6 ] 9.201645e+12 | [ 6 ] 9.300073e+12 | [ 6 ] 9.373574e+12 |
| [ 7 ] 9.385968e+12 | [ 7 ] 9.217297e+12 | [ 7 ] 9.176478e+12 | [ 7 ] 9.385776e+12 |
| [ 8 ] 9.385968e+12 | [ 8 ] 9.244814e+12 | [ 8 ] 9.244814e+12 | [ 8 ] 9.35861e+12 |
| [ 9 ] 9.385968e+12 | [ 9 ] 9.223724e+12 | [ 9 ] 9.218272e+12 | [ 9 ] 9.289174e+12 |
| [10] 9.385968e+12 | [10] 9.259906e+12 | [10] 9.273601e+12 | [10] 9.347346e+12 |
| [11] 9.385968e+12 | [11] 9.218272e+12 | [11] 9.169836e+12 | [11] 9.281285e+12 |
| [12] 9.385968e+12 | [12] 9.169836e+12 | [12] 9.293273e+12 | [12] 9.373574e+12 |
| [13] 9.385968e+12 | [13] 9.244814e+12 | [13] 9.177509e+12 | [13] 9.385968e+12 |
| [14] 9.385968e+12 | [14] 9.201645e+12 | [14] 9.217297e+12 | [14] 9.306581e+12 |
| [15] 9.385968e+12 | [15] 9.150173e+12 | [15] 9.253299e+12 | [15] 9.298835e+12 |
| [16] 9.385968e+12 | [16] 9.295777e+12 | [16] 9.174472e+12 | [16] 9.373574e+12 |
| [17] 9.385968e+12 | [17] 9.263678e+12 | [17] 9.259906e+12 | [17] 9.288699e+12 |
| [18] 9.385968e+12 | [18] 9.218272e+12 | [18] 9.286464e+12 | [18] 9.288699e+12 |
| [19] 9.385968e+12 | [19] 9.15123e+12 | [19] 9.217297e+12 | [19] 9.298835e+12 |
| [20] 9.385968e+12 | [20] 9.217297e+12 | [20] 9.273601e+12 | [20] 9.275185e+12 |
| [21] 9.385968e+12 | [21] 9.273601e+12 | [21] 9.285999e+12 | [21] 9.298835e+12 |
| [22] 9.385968e+12 | [22] 9.174472e+12 | [22] 9.176478e+12 | [22] 9.275185e+12 |
| [23] 9.385968e+12 | [23] 9.169836e+12 | [23] 9.244814e+12 | [23] 9.272078e+12 |
| [24] 9.385968e+12 | [24] 9.176478e+12 | [24] 9.263678e+12 | [24] 9.289174e+12 |
| [25] 9.385968e+12 | [25] 9.218272e+12 | [25] 9.297655e+12 | [25] 9.298835e+12 |
| [26] 9.385968e+12 | [26] 9.223724e+12 | [26] 9.291503e+12 | [26] 9.289174e+12 |
| [27] 9.385968e+12 | [27] 9.176478e+12 | [27] 9.292073e+12 | [27] 9.275185e+12 |
| [28] 9.391303e+12 | [28] 9.225762e+12 | [28] 9.217297e+12 | [28] 9.291894e+12 |
| [29] 9.385968e+12 | [29] 9.391291e+12 | [29] 9.174472e+12 | [29] 9.299075e+12 |
| [30] 9.385968e+12 | [30] 9.241439e+12 | [30] 9.27003e+12 | [30] 9.299075e+12 |

i=1,…, n where d >= 1,…, is the number of dimensions. Additionally, let K>=2 where K must be an integer and K={1,..,k}. For each group k; P={G(1),..,G(k)} of N, let μi be the centroid of the group G(i) for i ∈ K.

The principal reason for forming these groups is to define clusters for the shopkeeper dataset, in such a way that the total intra-cluster distance is minimized. This is achieved by developing the experimental categorization process using variants from the K-means algorithm.

We can observe this in figure 2:

1. We use the data set that will be processed.
2. A variant is selected, and the data is processed.
3. When the algorithm converges, the sum of the intra-cluster distance is stored.
4. Steps a, b and c are repeated 30 times.

5. The average for the intra-cluster distance is calculated, applying the selected variant.

In Table 2, we present the results from each of our thirty implementations, as well as the average intra-cluster distance. Importantly, if the average or all the interactions is considered, the variant that reports the shortest distance is the Forgy.

In all cases the Hartigan-Wong is the one that shows the longest intra-cluster distance. Among the most important results shown for the four variants is the fact that the Hartigan-Wong variant is much better for obtaining the longest intra-cluster distance and on several occasions both Lloyd and Forgy algorithms revealed the shortest intra-cluster distance (15 times).

However, most often it was the Lloyd algorithm that proved optimum for revealing the shortest intra-cluster distance.

## 5 Conclusions and Future Projects

In this article, we propose an amplification of a CMS such as Magento that will make it possible to incorporate machine learning functions and with this enable an exploratory analysis concerning the behavior of the main variants to the K-means algorithms, applied to a real set of data pertaining to shopkeepers in Mexico City.

In future work, we intend to combine these variants related to clustering algorithms with the proposal we discuss in the shopkeeper categorization phase. With the objective of categorizing the shopkeepers, we provide a ranking of products according to category.

This will inform us in terms of how products compete with each other, and thus show to what extent, the sale of products from different categories of shopkeepers is homogeneous (or heterogeneous). In today's competitive business environment, this will fulfill the great need that grocery wholesalers have to amass, monitor and analyze the data generated by their shopkeepers and competitors.

We are in an era when business represents a war, and like in any war, survival depends on the ability to act quickly in a changing environment. In this sense, we can affirm that the new wave of smart tools will be focused on tracking variables such as operational development, market conditions and the development of competitors, all of them in real time.

However, for many years, the traditional channel has been, as its name indicates, traditional. Despite being the favorite shopping approach throughout Latin America; its processes, mechanisms and tools have been at a standstill for decades. However, due to the pandemic, the digitalization of the corner shop has represented a major challenge for micro-enrepreneurs, in order to adapt to the new requirements of consumers. Smartphones offer a great advantage because they represent the gateway to digitalization.

In our experience, we have seen that some shopkeepers have doubts about how to use tools that are on their smartphones. Often, they do not know how to scan a code or they do not have an email address; these are minor issues but have great impact as it is precisely because of these details that they cannot access digital products and tools that would help them grow their business. If we are proposing the digitalization of small stores, then the discussion of digital education and training for the use of these tools is essential.

Although there have been real advances towards the digitalization of small stores, great challenges still exist. These businesses are fundamental to daily life; they are supply centers with great tradition and importance to the community. However, for them to continue operating in the long term and to capitalize on real growth opportunities, their gradual digital transformation will be essential.

## References

1. Diaz, R., Montalvo, R. F.: Innovative digital transformation strategies of large suppliers for mexican corner stores during a pandemic: challenges and opportunities. Handbook of Research on Digital Innovation and Networking in Post-COVID-19 Organizations, IGI Global, pp. 208–231 (2022) doi: 10.4018/978-1-6684-6762-6.ch011
2. Ziyadin, S., Suieubayeva, S., Utegenova, A.: Digital transformation in business. Lecture Notes in Networks and Systems, Springer International Publishing, pp. 408–415 (2019) doi: 10.1007/978-3-030-27015-5_49
3. Palmatier, R. W., Crecelius, A. T.: The "first principles" of marketing strategy. Academy of Marketing Science Review, vol. 9, no. 1–2, pp. 5–26 (2019) doi: 10.1007/s13162-019-00134- y
4. Kumar, S. J., Oommen-Philip, A.: Achieving market segmentation from B2B insurance client data using RFM and k-means algorithm. In: IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems, pp. 463–469 (2022) doi: 10.11 09/spices52834.2022.9774051
5. Thomas, J., Preethi, N.: Customer segmentation in the field of marketing. In: 4th International Conference on Recent Trends in Computer Science and Technology, pp. 401–405 (2022) doi: 10.1109/icrtcst54752.2022.9781964
6. Arora, P., Deepali, D., Varshney, S.: Analysis of K-means and K-medoids algorithm for big data. Procedia Computer Science, vol. 78, pp. 507–512 (2016) doi: 10.1016/j.procs.20 16.02.095
7. Dogan, O., Ayçin, E., Bulut, Z. A.: Customer segmentation by using RFM model and clustering methods: a case study in retail industry. International Journal of Contemporary Economics and Administrative Sciences, vol. 8, no. 1, pp. 1–19 (2018)
8. Tavakoli, M., Molavi, M., Masoumi, V., Mobini, M., Etemad, S., Rahmani, R.: Customer segmentation and strategy development based on user behavior analysis, RFM Model and data mining techniques: A case study. In: IEEE 15th International Conference on e-Business Engineering (2018) doi: 10.1109/icebe.2018.00027
9. Anitha, P., Patil, M. M.: RFM model for customer purchase behavior using K-means algorithm. Journal of King Saud University - Computer and Information Sciences, vol. 34, no. 5, pp. 1785–1792 (2022) doi: 10.1016/j.jksuci.2019.12.011
10. Wu, J., Shi, L., Lin, W. P., Tsai, S. B., Li, Y., Yang, L., Xu, G.: An empirical study on customer segmentation by purchase behaviors using a RFM model and K-means Algorithm. Mathematical Problems in Engineering, vol. 2020, Hindawi Limited, pp. 1–7 (2020) doi: 10.1155/2020/8884227
11. Das, S.: A systematic study of integrated marketing communication and content management system for millennial consumers. Innovations in Digital Branding and Content Marketing. IGI Global, pp. 91–112 (2021) doi: 10.4018/978-1-7998-4420-4.ch005
12. Howe-Patterson, K., Schuiling, I.: Shopify in Germany: An analysis of a Canadian e-commerce platform's marketing strategy and activities in an international market (2020)

13. Sims, L.: Building your online store with WordPress and WooCommerce. Apress (2018) doi: 10.1007/978-1-4842-3846-2
14. Tsagkias, M., King, T. H., Kallumadi, S., Murdock, V., de Rijke, M.: Challenges and research opportunities in eCommerce search and recommendations. Association for Computing Machinery, vol. 54, no. 1, pp. 1–23 (2020) doi: 10.1145/3451964.3451966
15. Abdullah, E. N., Ahmad, S., Ismail, M., Diah, N. M.: Evaluating e-Commerce website content management system in assisting usability issues. In: IEEE Symposium on Industrial Electronics and Applications (2021) doi: 10.1109/isiea51897.2021.9509991
16. Scott, C., Fehlig, V.: Small retailers in/out of the crisis in Germany – Re-inventing retailing. In: SHS Web of Conferences, EDP Sciences, vol. 116., p. 00052 (2021) doi: 10.1051/shsconf/202111600052
17. Hensel, R., Visser, R., Overdiek, A., Sjoer, E.: A small independent retailer's performance: Influenced by innovative strategic decision-making skills? Journal of Innovation and Knowledge, vol. 6, no. 4, pp. 280–289 (2021) doi: 10.1016/j.jik.2021.10.002
18. Liu, J., Liao, X., Huang, W., Liao, X.: Market segmentation: A multiple criteria approach combining preference analysis and segmentation decision. Omega, vol. 83, pp. 1–13 (2019) doi: 10.1016/j.omega.2018.01.008
19. Xian, Z., Keikhosrokiani, P., XinYing, C., Li, Z.: An RFM model using K-means clustering to improve customer segmentation and product recommendation. Advances in Marketing, Customer Relationship Management, and E-Services, IGI Global, pp. 124–145 (2022) doi: 10.4018/978-1-6684-4168-8.ch006
20. Shirazy, A., Hezarkhani, A., Shirazi, A., Khakmardan, S., Rooki, R.: K-means clustering and general regression neural network methods for copper mineralization probability in Char-Farsakh, Iran. Türkiye Jeoloji Bülteni / Geological Bulletin of Turkey, vol. 65, no. 1, pp. 79-92 (2021) doi: 10.25288/tjb.1010636
21. Mittal, M., Sharma, R. K., Singh, V. P.: Performance evaluation of threshold-based and K-means clustering algorithms using iris dataset. Recent Patents on Engineering, vol. 13, no. 2, pp. 131–135 (2019) doi: 10.2174/1872212112666180510153006
22. Fränti, P., Sieranoja, S.: K-means properties on six clustering benchmark datasets. Applied Intelligence, vol. 48, no. 12, pp. 4743–4759 (2018) doi: 10.1007/s10489-018-1238-7
23. Fränti, P., Sieranoja, S.: How much can K-means be improved by using better initialization and repeats? Pattern Recognition, vol. 93, pp. 95–112 (2019) doi: 10.1016/j.patcog.2019.04.014
24. Sreevalsan-Nair, J.: K-Means clustering. Encyclopedia of mathematical geosciences, Springer International Publishing, pp. 1–3 (2021) doi: 10.1007/978-3-030-26050-7_171-1
25. Almanza-Ortega, N. N., Pérez-Ortega, J., Zavala-Díaz, J. C., Solís-Romero, J.: Comparative analysis of K-means variants implemented in R. Computación y Sistemas, vol. 26, no. 1 (2022) doi: 10.13053/cys-26-1-4158
26. Wu, B.: K-means clustering algorithm and python implementation. In: IEEE International Conference on Computer Science, Artificial Intelligence and Electronic Engineering (2021) doi: 10.1109/csaiee54046.2021.9543260
27. Thrun, M. C., Ultsch, A.: Using projection-based clustering to find distance- and density-based clusters in high-dimensional data. Journal of Classification, vol. 38, no. 2, pp. 280–312 (2020) doi: 10.1007/s00357-020-09373-2
28. Thrun, M. C., Ultsch, A.: Clustering benchmark datasets exploiting the fundamental clustering problems. Data in Brief, vol. 30, pp. 105501 (2020) doi: 10.1016/j.dib.2020.105501
29. Guo, G., Altrjman, C.: E-Commerce customer segmentation method under improved K-Means Algorithm. In International Conference on Multi-modal Information Analytics, Springer, Cham, vol. 138, pp. 1083–1089 (2022) doi: 10.1007/978-3-031-05484-6_148