

Market segmentation: Machine Learning in Marketing in the Context of COVID-19

PEDRO PABLO CHAMBI CONDORI ¹

SUBMITTED: 15/09/2022 ACCEPTED: 14/02/2023 PUBLISHED: 31/07/2023

ABSTRACT

The COVID-19 health crisis has led to unprecedented changes in consumer behavior, as consumers now purchase differently and use different means. Consumers are checking and judging products via electronic devices, shaping trends in consumer segments. This research study aimed to use the clustering model with Machine Learning resources in the analysis of clusters as a resource for consumer segmentation, a major component in business marketing management. A 6-question questionnaire was administered to 506 people ranging from 18 to 65 years old to gauge their opinions about going shopping. A dataset was organized using the data collected and processed using RapidMiner Studio 9.10 software. The optimal number of clusters and their components were obtained from the performance indicator provided by Machine Learning.

Keywords: market research; segmentation; artificial intelligence; COVID-19.

INTRODUCTION

Clustering is a set of non-inferential descriptive statistical techniques—unsupervised in machine learning— aimed at grouping similar objects in the same cluster, by which clusters with high internal affinity and external difference are formed. In business marketing, market segmentation provides insight into the composition of the market, differentiating groups with similar characteristics and needs. Customer segmentation is a fundamental tool for a company, as it allows for the identification of common and differentiated characteristics within the customer database, which are very useful for the strategic management of customers.

Casco (2020) studied the impact of COVID-19 on consumer behavior, finding that consumers are increasingly using technology to purchase goods and services, as well as delivery services for the delivery of products. Mehta et al. (2020) also studied consumer behavior in the context of COVID-19 and found differences in consumer behavior in ordinary times and in times of crisis; they also identified changing needs due to cultural factors and the age of customers, which have an impact on new patterns of consumer behavior. Eger et al. (2021) studied the trends and impact of COVID-19 on consumer behavior, documenting behavioral patterns over the course of the second wave of the pandemic in the Czech Republic. They found intergenerational similarities and differences.

The coronavirus outbreak proved to be a devastating human tragedy with negative impacts on the global economy that altered marketing concepts, the very context of marketing tactics, and marketing strategies (Kumar et al., 2020). This unprecedented crisis forced people to isolate themselves, disrupting all aspects of their regular activities (Kabir & Madria, 2021). The pandemic has wreaked havoc, and faced with this situation, many researchers from different fields combined their efforts to provide a wide range of solutions, including advancing digital marketing (Rodríguez-Rodríguez et al., 2021). Yıldırım et al. (2021) studied the negative impacts of COVID-19 and found that vulnerability,

¹ PhD. Associate professor at the School of Commercial Engineering of Universidad Nacional Jorge Basadre Grohmann de Tacna (Tacna, Peru).
Orcid: <https://orcid.org/0000-0002-8628-6825>
E-mail: pchambic@unibg.edu.pe

perceived risk and fear can significantly increase preventive behaviors in consumers, providing evidence of dynamic variation in consumer clustering, which are to be studied through market and customer segmentation for strategic customer management.

The impact of 4th generation technologies on market management, quality configuration, product design and development, and communication with customers is becoming increasingly significant. In this regard, Brei (2020) states that machine learning offers many potential applications in marketing. That is, marketing can be transformed into a more autonomous scientific work through the use of data and the proper formulation of each application. There are two main traditional marketing paradigms: destination marketing and relationship marketing. Each of the many aspects of the two marketing paradigms can be expressed as a machine learning problem, i.e., for each problem, a machine learning model can be built, and model parameters can be set. Artificial intelligence expressions have merited studies because machines with deep learning capabilities can take digital marketing to higher levels and make a difference (Miklosik et al., 2019). Lately, digital marketing is in a state of constant evolution, introducing new tools for data processing that describe the new consumer habits, where the success or failure of digital communication depends largely on the quality of its content marketing (Baltes, 2015).

The widespread impacts of artificial intelligence (AI) and machine learning (ML) on many segments of society has been strongly felt in the field of marketing (Jarek & Mazurek, 2019). After all, machine learning offers a variety of benefits, including the opportunity to use proven techniques for the generalization of the advances achieved by science and scientific research.

According to Ullal et al. (2021), machine learning is a technique that combines science, statistics and computational coding to detect patterns in large volumes of data. Based on this data support, it is possible to predict future estimates (expected sales, market segment attention quota, and customer management under the CRM concept) resulting from the application of mathematical and statistical algorithms. The integration of big data and machine learning techniques is a result of three key factors: the availability of algorithms and data, the increased processing power of computers and the lower cost of storage of digital devices. The speed of developments in computer science and software engineering, however, poses a challenge of ever-increasing importance: identify

and implement the right combination of hardware platforms and software architectures to ensure the continuity of the integrated model and minimize its obsolescence, mitigating a loss of the processing capacity of the data management platform.

The purpose of this study is to conduct a case study on the application of one of the resources of artificial intelligence in market research, specifically, customer segmentation as a tool for the strategic management of business marketing.

The clustering model was used in this study; the rationale of cluster analysis is to group objects into clusters using clustering algorithms. K-means is the most popular of the clustering algorithms, which groups numerical data and each cluster has a center called the mean. It is assumed that the number of clusters k is a fixed value.

According to Mahendiran et al. (2012), the cited k -means algorithm consists of the squared Euclidean distance divided by two variables, in other words, the sum of squares of the differences of all the coordinates of two points as shown in the following equation:

$$d(i, j) = \sqrt{\sum_k (X_{ik} - X_{jk})^2}$$

Where X_{ik} and X_{jk} are the individual values of each of the points, and the Davies-Bouldin index, which calculates the distances between clusters, is used to evaluate performance. In their study *Estimating Marketing Component Effects: Double Machine Learning from Targeted Digital Promotions*, Ellickson et al. (2022, p. 2) first aimed to demonstrate how machine learning can be leveraged in a unified framework to provide causal estimates of the effect of separate components in a high-dimensional marketing intervention, using only observational data. Next, the methodology was applied to data from 34 distinct e-mail promotions sent by a retailer to more than 1.3 million people on its contact list to highlight quantitative findings that are relevant to both the marketing literature and management practice. Furthermore, Dzyabura and Hauser (2011) assert that heuristic decision rules predict validation decisions better than compensatory rules.

Jordan and Mitchell (2015), on the other hand, assert that machine learning assists computer manufacturers in building intelligent equipment that improves through experience. Machine learning has experienced a significant increase in recent years, developing at the intersection of computer science

and quantitative methods, and at the core of artificial intelligence and data processing. The accelerated advancement in machine learning has been driven both by the rapid incursion of innovative algorithms, programming languages, increased processor capacity and the expansion of learning theories, and the steady growth in the accessibility of online data, data repositories, and available low-cost computing resources. The adoption of data-intensive machine learning methods is evident throughout the sciences, business management, commerce, finance and economics, health management, education, machine vision surveillance, and engineering, leading to improved decision quality based on data analytics and providing decision-makers with evidence for quality decision-making.

According to Jain and Kain (2018, p. 131), Machine Learning is the science of getting computers to operate automatically. In the past decade, machine learning has given us self-driving cars, voice recognition apps, effective web search, and an improved understanding of the human genome. Their research paper discussed the method to predict stock market behavior using machine learning algorithms: linear regression, random forest, and multilayer perceptron.

Abakouy et al. (2019, p. 1) conducted a comparative study on the most popular machine learning methods applied to the challenging problem of e-mail marketing personalization. Subject and sender lines have a strong influence on click rates of e-mails, as customers often open and click on emails based on the subject and sender. According to Kaličanin et al. (2019, p. 472), artificial intelligence (AI) impacts many aspects of life through smart devices and applications, designed to understand consumer behavior, needs, and preferences in order to deliver personalized experiences. AI has been one of the primary drivers of innovation in marketing. Marketers are already leveraging the advantages of AI to gain valuable insights into customers, competitors, and markets. AI automates tasks, reduces costs, and improves workflows. Their paper examines the current and potential applications of AI within marketing by providing a comprehensive overview of existing academic research.

On the marketing intelligent approach, Mahajan et al. (2017, p. 21929) state that AI is a state-of-the-art marketing management support system for dealing with knowledge using machine learning and other soft computing techniques. The range of potential applications of machine learning techniques in marketing management are consumer behavior, product market structure optimization, marketing mix

management, strategic marketing, and finance domain, among others. They describe the synergy between marketing and intelligent systems, especially machine learning techniques.

Regarding the evolution of digital marketing, which was further leveraged in the COVID-19 health crisis context, Bayoude et al. (2018, p. 373) argued that digital marketing is constantly evolving, new tools are regularly introduced in line with new consumer habits and the multiplication of data, often forcing marketers to delve into too much data that may not even provide them with the overview they need to make business decisions. Following the revolution of machine learning technology in other real-world applications, machine learning is changing the digital marketing landscape. In 2018, marketing organizations are implementing or expanding their use of machine learning. It becomes easier to predict and analyze consumer behavior with great accuracy. In their paper, they proposed the use of state-of-the-art and the most commonly used potential machine learning models in various digital marketing strategies. Also they showed how machine learning tools can be used on a large scale for marketing purposes by analyzing extremely large data sets. The way ML is integrated into digital marketing practices helps them to better understand target consumers and optimize their interactions with them.

The application of artificial intelligence (AI) in marketing aims to continuously track and forecast the next purchasing decisions of target consumers and improve their consumer "journey". In this regard, MR (2021) states that the potential of AI is reflected in its core elements: big data, machine learning and powerful solutions. Big data enables marketers to aggregate and segment large amounts of data with minimal manual work. Using such data, they will be confident that they will deliver the right message to the right individuals in the right circumstances, through the channel of their choice. Machine learning (deep learning) enables marketers to understand and draw logical conclusions from large collections of data. They can predict consumer trends, track and quantitatively analyze consumer purchases, forecast consumer behavior, and predict the next consumer behavior. We live in an era where machines truly understand the world in the same way that humans do. Machines can easily recognize concepts and themes in a variety of data, interpret human emotions and communications, and generate appropriate responses to consumers. They can easily predict buyer behavior and decisions and use that data to solve problems in the future. In the coming years, marketers can

expect a greater impact of AI, through more intelligent search, smarter ads, refined content delivery, reliance on bots, continuous learning, fraud and data breach prevention, feeling analysis, image and voice recognition, sales forecasting, language recognition, predictive customer service, customer segmentation, and more (MR, 2021). In addition, they may discover a stronger future relationship between marketers and artificial intelligence machines.

Prompted by the COVID-19 health crisis, it is undeniable that the world is rapidly shifting to a digital age. People consume more digital content on a daily basis as they spend more time online. Digital tools and the sites they use play an increasingly significant role in their lives. Smart marketing companies recognize this fact and incorporate the concept of artificial intelligence into their marketing strategies. Thiraviyam (2018) claims that artificial intelligence is revolutionizing the marketing performance universe today. Big data and the exponential growth of computing power have paved the way for the takeoff of artificial intelligence. As advances continue to be made in machine learning, neural networks and deep learning technology, more companies are turning to artificial intelligence to make their operations smarter and more efficient. The marketing world is increasingly embracing these resources to make sense of data, learn more about customers, and optimize operations. This paper introduces artificial intelligence marketing and further explores how marketers are harnessing the power of artificial intelligence and discusses the need to use this strategy for marketing products and services.

Machine learning offers great potential value for marketing-related applications. However, the proliferation of data types, methods, tools, and programming languages hinders knowledge integration among marketing analytics teams, making collaboration difficult. In this regard, Villarroel and Silipo (2021, p. 393) state that visual programming might facilitate the orchestration of ML projects in a more intuitive visual fashion. In terms of marketing strategy, Huang and Rust (2022, p. 218) state that mechanical artificial intelligence can be used for segmentation (segment recognition), thinking artificial intelligence for targeting (segment recommendation), and feeling artificial intelligence for positioning (segment resonance).

The literature review above explains that market and customer segmentation is a fundamental task for the management of business marketing because it relates to the identification of groups of consumers who interact with the products

produced by a company. Market research involves gathering consumer perception, often using statistical resources that provide forecasts with a level of probability of success and a degree of statistical significance; however, artificial intelligence resources provide statistical modeling with a degree of certainty in the forecasts.

Machine Learning modeling uses mathematical, econometric, and statistical models. In this case, the k -means model was used, which is a quantitative tool used for clustering that divides the universe n of observations into k different groups guided by criteria of internal affinity and external difference. Figure 1 illustrates the k -means model obtained using Jupyter Notebook and Anaconda 3.0, which seeks to include the nearest neighbor in a specific group. This algorithm falls into the group of unsupervised algorithms. Clustering of the n observations into the k different groups is performed by minimizing the sum of distances in each observation and the centroid of the cluster.

Depending on the study's objectives and the analysis of the cases of identification of cluster groups, the segmentation processes with the k -means algorithm aim to minimize the variance within a cluster, referred to as the search for internal homogeneity. On the one hand, it is essential to find homogeneous groups that can be satisfied with a given marketing mix, and, on the other hand, a wide external heterogeneity between the different groups of a population is desired. Therefore, the selection of higher variances between groups of clusters will be favored. It is observed in Figure 1 that there are 5 groups and each of them has components that make up the cluster. Cluster analysis is critical in marketing because its purpose is to identify consumer behavior, find opportunities for new products, choose test markets, and decant data.

Figure 2 shows the path of cluster formation according to the distance measure used. The number of clusters depends on where the dendrogram is cut, thus the decision on the optimal number of clusters is subjective. It is advisable to choose an interpretable number of clusters, based on the number of clusters desired, e.g. high-performance clusters, high-availability clusters, and high-efficiency clusters.

This study analyzes the implications of the adoption of artificial intelligence technologies for the location of better market and customer segmentation models, which are particularly important for companies engaged in market and customer management. The quantitative models that describe the behavior

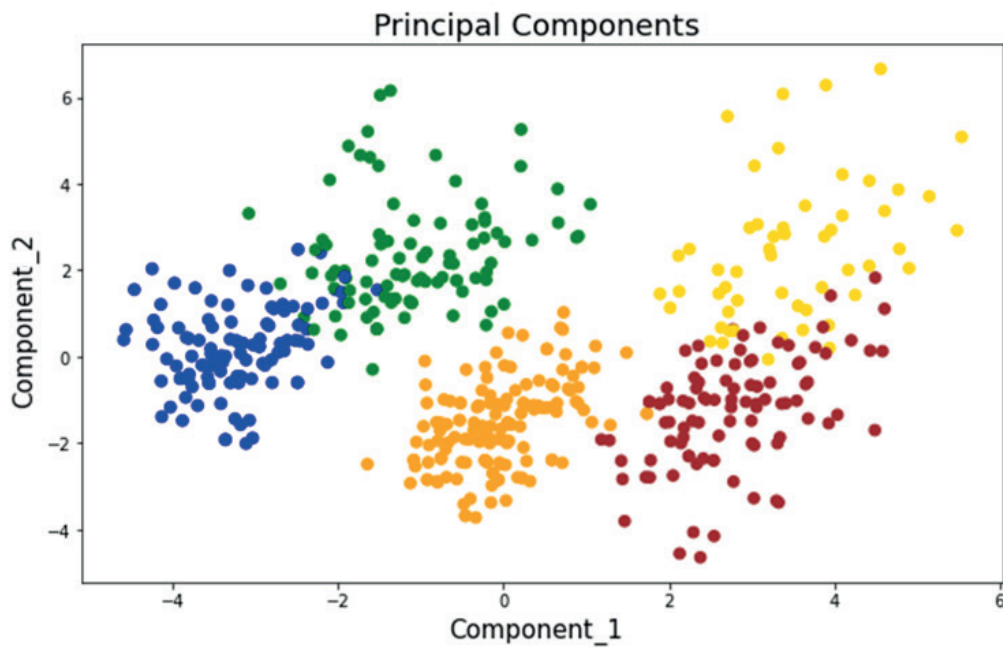


Figure 1. K-Means Clustering.

Source: Prepared by the author.

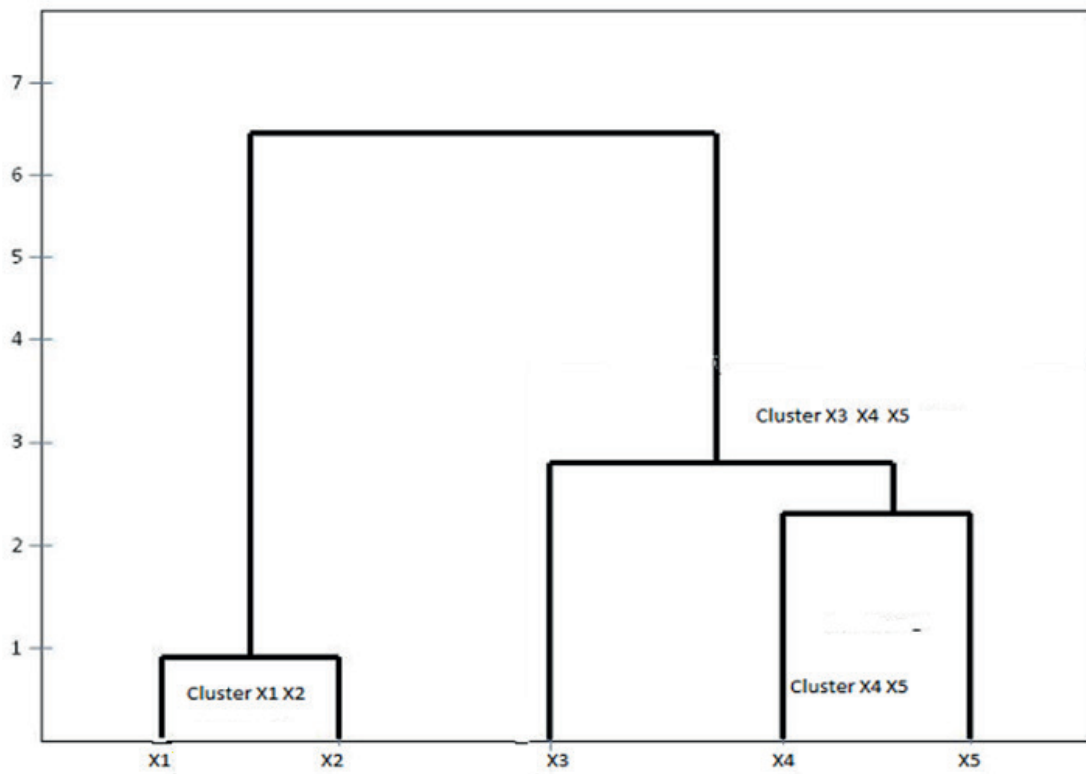


Figure 2. Graphical Representation of Hierarchical Segmentation (Dendrogram).

Source: Prepared by the author using Anaconda 3.0.

of people in the purchasing decision process are multivariate, i.e. a variety of qualitative factors are involved in people's purchasing decisions. A major factor in the purchasing decisions made by individuals is the set of tastes, preferences, and lifestyles of individuals. The distribution of such manifestations and the advancement of technology have allowed consumers to have access to information almost instantaneously. As a result, small segments of consumer groups with specific consumer profiles have emerged and their requirements prior to making a decision to purchase a given product are considered.

This research reveals the importance of adopting artificial intelligence models (machine learning) in the implementation of segmentation models applied by companies for the strategic management of their customers and markets. It is all the more important in these times of COVID-19 to permanently evaluate the dynamics of changes in consumers' tastes and preferences as it allows us to assess the efficiency of the application through a performance measurement indicator, namely the Davies-Bouldin index for this study. Clustering models implemented with artificial intelligence tools can be applied in all business and social environments to identify highly differentiated groups that show significant attributes of internal cohesion. Results can be used to implement dynamic, proactive, and innovative policies in organizational management.

From the above, the research hypothesis is as follows: The adoption of artificial intelligence technologies has a positive influence on customer segmentation practices in market research validated by the degree of certainty in segmentation results.

METHODOLOGY

This is an analytical study with a non-experimental design. A 6-question questionnaire with a scale of 7 levels of responses, where only one alternative could be selected for each question, was elaborated. In order to obtain information on the perception of

shopping, the questionnaire was addressed to the public that goes to shopping malls to do their shopping. Six (6) attitudinal variables were identified for this purpose. Consumers were asked to express their degree of agreement with the statements described below:

- V1: Shopping is fun.
- V2: Shopping is not good for my budget.
- V3: When I go shopping, I take the opportunity to eat out.
- V4: When I go shopping, I look for the best deals.
- V5: I don't feel motivated to go shopping.
- V6: I can save a lot of money when I have the chance to shop around.

Each of the questions had seven levels of response, where the lowest level was 1 (totally disagree), and the highest level was 7 (totally agree).

As a matter of convenience, the sample consisted of 506 people of both sexes over 18 years of age; the sample is not random. The questionnaire was administered through Google Forms and was available online for 15 days during March 2022. The Guttman test was selected for the reliability analysis with an index of 0.659 and the validity index with Bartlett's test of sphericity using the X^2 indicator with a p -value < 0.05 at 95% statistical significance.

Upon collecting the data using the questionnaire format, the dataset containing the collected data was prepared and organized in an Excel sheet and then exposed to the clustering model with RapidMiner Studio 9.10, following the process presented in Figure 3. The segmentation tests were executed applying the k-means unsupervised learning model. First, the data of 506 records were uploaded, then the data was subjected to the normalization process, then the clustering was activated and, finally, the results were subjected to the performance evaluation.



Figure 3. Cluster Modeling with RapidMiner Studio 9.10

Source: Prepared by the author.

Figure 3 describes the sequential process of clustering application. The database is imported into the Rapid Miner environment; afterward, each of the variables is normalized because the ML modeling requires data that have a Gaussian distribution; next, the cluster classification model is applied to obtain the results for different groups searching for the smallest or largest distance between groups of people; lastly, the results are subjected to the performance evaluation, finding the optimal number of groups that have the best performance.

As presented in Figure 3, the method used for cluster execution with RapidMiner 9. 10 consists of the squared Euclidean distance divided by two variables, in other words, the sum of squares of the differences of all the coordinates of two points as shown in the following equation:

$$d(i, j) = \sqrt{\sum_k (X_{ik} - X_{jk})^2},$$

Where X_{ik} and X_{jk} are the individual values of each of the points, and the Davies-Bouldin index is used

to evaluate performance. The Davies-Bouldin model is a resource used for the internal validation of clustering, which is expressed as follows:

$$DB = \left(\frac{1}{K}\right) \sum_{i=1}^k \max\left(\frac{\sigma_i + \sigma_j}{d(C_i, C_j)}\right)$$

Where k is the number of clusters, σ_i is the average distance between each point in cluster i and the centroid of the cluster, σ_j is the average distance between each point in cluster j and the centroid of the cluster, and $d(C_i, C_j)$ is the distance between the cluster centroids.

K-means clustering algorithm is an unsupervised clustering algorithm that has high scalability for processing data. The number of groups to be formed must be specified to use k -means; it is possible to find the optimal number of k clusters by testing, which is the main objective of the research.

RESULTS

Figure 4 depicts the descriptive behavior of each of the variables. Respondents' responses were mostly

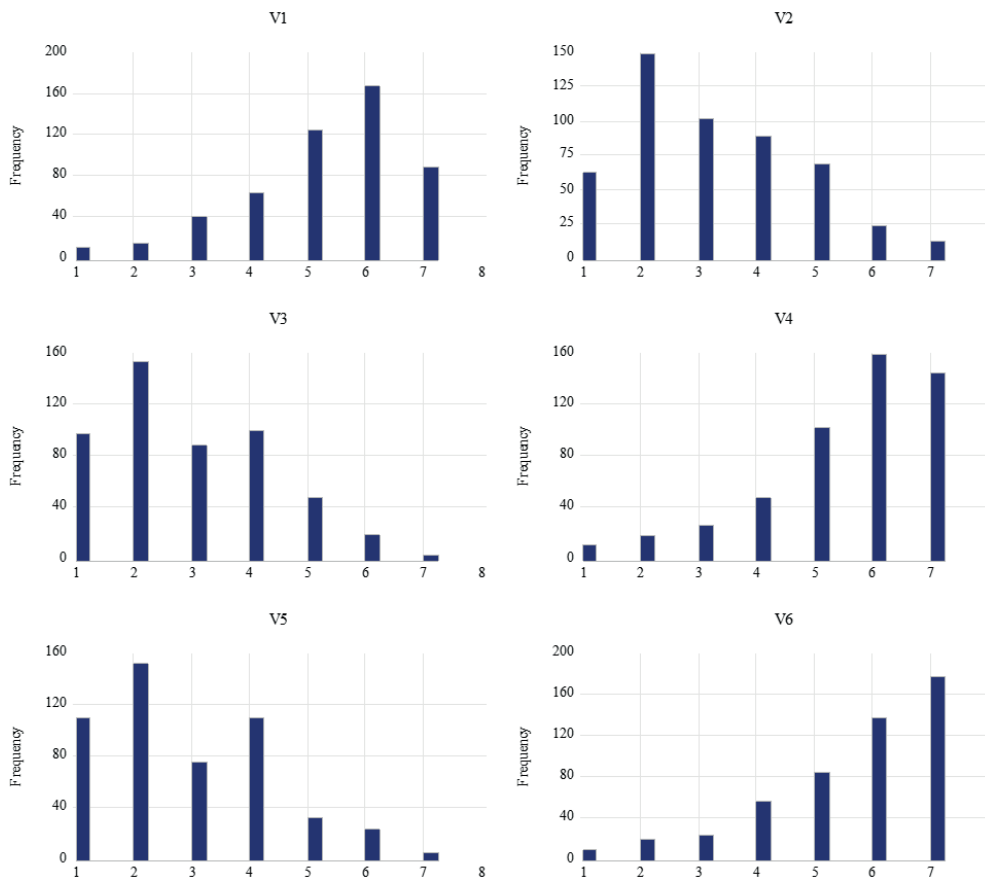


Figure 4. Distribution of Responses to Segmentation Statements.

Source: Prepared by the author based on data from the questionnaire.

positive for all the statements; furthermore, respondents find shopping to be a fun; an opportunity to price-shop; an opportunity to take advantage of bargains; and a chance to find savings in their personal budgets.

Table 1 shows the descriptive indicators obtained from the questionnaire data, which refer to the distribution of the answers given by the respondents to each statement. Among the 6 variables, V5 (I don't feel motivated to go shopping) has the lowest average value. As can be observed, it is a negative statement to which respondents gave the lowest value; therefore, the answers provided are positive. Conversely, a higher positive value corresponds to V6 (I can save a lot of money when I have the chance to shop around), as it is only natural that respondents value having as much information as possible to make a purchase decision.

In the correlation matrix in Table 2, direct relationships are observed between variables V1, V4, and V6 consistent with the descriptive illustration of the distributions shown in Figure 4. In turn, these

three variables have indirect relationships with variables V2, V3, and V5. In the matrix, the highest correlations are identified, namely the relationship between variables V4 and V6 with $r = 0.5721$, and also the lowest correlations, such as the one between variables V1 and V5 with $r = -0.3451$. Such information is very important for the formation of clusters in the process of group construction.

The centroid indicator between clusters is the distance between clusters. The centroids shown in Table 3 were obtained using segmentation with $k = 1$, $k = 2$, $k = 3$, $k = 4$, and $k = 5$. Centroids with high values represent the heterogeneity between the clusters formed via segmentation. Considering the high value of the centroids, the test with $k = 2$ exhibits pronounced heterogeneity with respect to the other groups. Regarding the variables, those with lower centroids are the variables V2, V3, and V5, whereas V1, V4, and V6 exhibit large centroids among the groups formed.

Table 4 shows the summary of the analysis of variance using the Univariate-F test for each of the

Table 1. Descriptive Indicators of the Sample.

Indicator	V1	V2	V3	V4	V5	V6
Mean	5.2253	3.1482	2.8320	5.4921	2.7846	5.5850
Median	6.0000	3.0000	3.0000	6.0000	2.0000	6.0000
Maximum	7.0000	7.0000	7.0000	7.0000	7.0000	7.0000
Minimum	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Std. Dev.	1.4144	1.5142	1.4300	1.4762	1.4702	1.4971
Skewness	-0.8732	0.5118	0.5245	-1.1214	0.6147	-1.0996
Kurtosis	3.4038	2.5153	2.4708	3.8242	2.6413	3.6224
Jarque-Bera	67.7465	27.0447	29.1061	120.3733	34.5784	3.6224
Probability	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Observations	506	506	506	506	506	506

Source: Prepared by the author based on the data from the questionnaire.

Table 2. Unconditional Correlation Matrix.

Question	V1	V2	V3	V4	V5	V6
V1	1.0000	-0.0711	0.0657	0.4257	-0.3451	0.3935
V2	-0.0711	1.0000	0.0408	0.0081	0.2883	0.0571
V3	0.0657	0.0408	1.0000	-0.0227	0.1297	0.0257
V4	0.4257	0.0081	-0.0227	1.0000	-0.1354	0.5721
V5	-0.3451	0.0081	0.1297	-0.1354	1.0000	-0.0496
V6	0.3935	0.2883	0.0256	0.5721	-0.0490	1.0000

Source: Prepared by the author based on the data from the questionnaire.

Table 3. Centroid Indicators of the Clusters Formed with the Sample.

Variable	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
V1	5.7566	3.9570	6.1053	3.4074	5.6022
V2	1.9934	4.2473	4.6579	2.7963	2.2903
V3	1.8421	3.0323	2.8947	2.3333	4.4624
V4	6.1184	5.2258	6.2456	2.7778	5.3871
V5	2.0461	4.7312	2.4035	2.3333	2.7742
V6	6.1579	5.5376	6.3158	2.9074	5.3548

Source: Prepared by the author based on the data from the questionnaire.

Table 4. Analysis of Variance (ANOVA).

Variable	Mean Square Cluster	df	Mean Square Error	df	F	Sig.
V1	76 457.00	5	1.256	500	60.871	.000
V2	128 826.00	5	1.028	500	125.376	.000
V3	105 626.00	5	1.009	500	104.665	.000
V4	116 132.00	5	1.040	500	111.706	.000
V5	100 875.00	5	1.174	500	85.903	.000
V6	105 712.00	5	1.213	500	87.180	.000

Source: Prepared by the author based on the data from the questionnaire.

variables included in the analysis. It is obtained by using the cluster groups as a factor and each of the variables as a dependent variable. The analysis of variance is conducted by taking into account the classification of the groups as a factor. In this case, the 5 groups become the factor levels, thus rejecting the equality of the groups' means, and accepting the existence of heterogeneity between groups and homogeneity within the groups. Therefore, the five organized groups were validated with the k -means algorithm using RapidMiner 9.10, confirming the statistical consistency of intra-cluster homogeneity and inter-cluster heterogeneity, given the p -value criterion at 95% confidence.

Table 5 shows the performance indicators for four tests. The cluster with the lowest DB value in the performance measurement of the model of the four tests corresponds to $k=4$, with a performance percentage of 24.30%, followed by $k=5$ with a performance of 25.40%. Therefore, considering the Davies-Bouldin index, the optimal number of clusters is $k=4$, with 196 components in group1, 56 in group 2, 133 in group 3, and 121 in group 4, totaling 506 respondents, as shown in Figure 5.

Bartlett's test has been used as an auxiliary tool to estimate the number of groups that can be formed using the data under study. This method uses the average distance of the observations to their centroid;

it focuses on the internal homogeneity of a group. The greater the number of groups (k), the more the internal variance of a cluster tends to decrease, i.e., the smaller the internal distance between the elements of a cluster, the better, since the groups are more internally homogeneous.

The Davis-Bouldin indicator (Table 5), which is used to perform the internal validation of the groups formed, is used to confirm the hypothesis testing. The lowest DB value obtained from among the four cluster options corresponds to the segment that has 24.30%. This is the optimal segmentation option, the degree of certainty of which is provided by machine learning, thus demonstrating the research hypothesis.

DISCUSSION

Research on the subject of consumer groups is particularly useful for a company that operates in various markets offering its products, not only for the strategic management of its products but also for the management of markets and customers. This activity, from the perspective of market research, focuses on gathering the perceptions of tastes and preferences of consumers in relation to certain products on the market. Provided with accurate information on consumer tastes and preferences, companies can quickly align their offerings through

Table 5. *K-Means Segmentation Efficiency Measurement.*

No. of Segments	Clusters	Components	Davies-Bouldin
2	Cluster 1	333	29.00%
	Cluster 2	173	
	Total:	506	
3	Cluster 1	212	27.90%
	Cluster 2	107	
	Cluster 3	187	
	Total:	506	
4	Cluster 1	196	24.30%
	Cluster 2	56	
	Cluster 3	133	
	Cluster 4	121	
	Total:	506	
5	Cluster 1	152	25.40%
	Cluster 2	93	
	Cluster 3	114	
	Cluster 4	54	
	Cluster 5	93	
	Total:	506	

Source: Prepared by the author based on the data from the questionnaire.

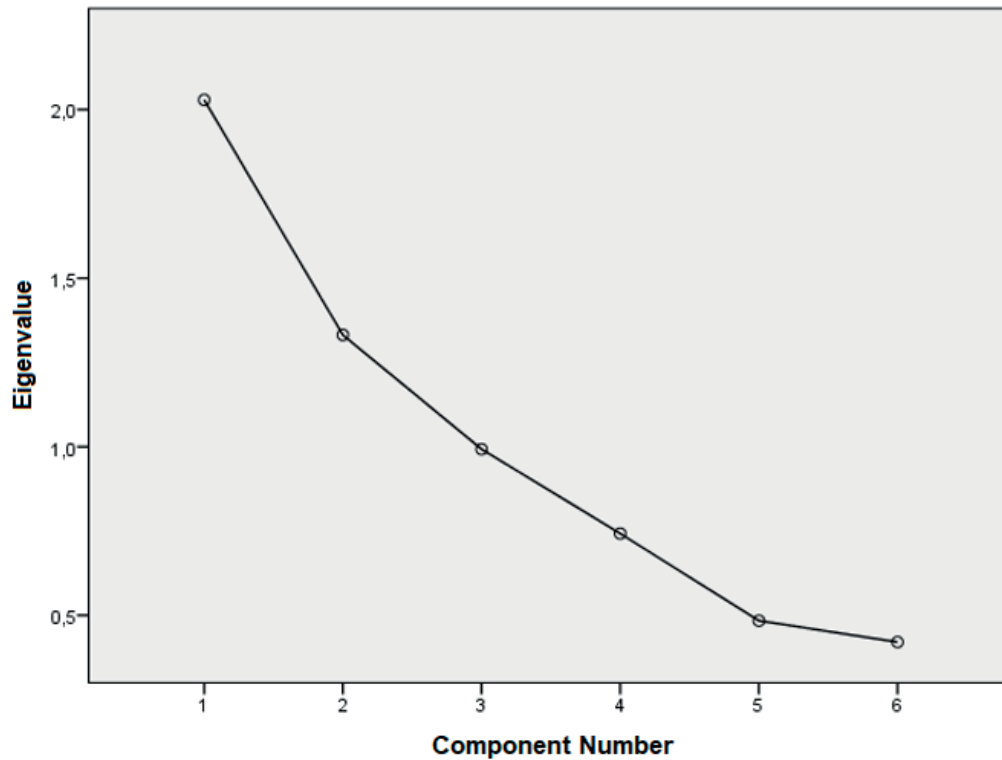


Figure 5. Analysis to Determine Number of Groups.

Source: Prepared by the author based on the data from the questionnaire.

their innovation departments, targeting certain products to specific consumer groups.

Consumer relationship management aims to build relationships by segmenting customers and designing appropriate marketing tools. Smart marketing companies recognize this and incorporate the concept of artificial intelligence into their marketing strategies. Currently, artificial intelligence is revolutionizing marketing (Thiraviyam, 2018).

Marketing science has a long tradition of accepting new challenges, new methods, and new disciplines. For example, the adoption of computational technologies for market research activities; the use of the clustering algorithm is another alternative application of artificial intelligence (AI) for the identification of consumer groups. Today, marketing science rests on the diverse efforts of researchers who, for nearly 50 years, have synthesized solutions from various disciplines to bring a fresh perspective to address marketing problems. More often than not, the melting pot of marketing science has provided better and more robust models and methods to other disciplines.

Recent progress in machine learning has been positively impacted by the accelerated development of new algorithms and learning theories as well as by the steady growth in the accessibility of online data, data repositories, and available low-cost computing resources. The adoption of data-intensive machine learning methods can be found in all expressions of science, technology, and commerce, which in the field of business marketing drives decision-makers to rely more on evidence than intuition (Jordan & Mitchell, 2015).

Marketing leverages the advantages of AI and machine learning models to automate, optimize, and transform data into actions and interactions to forecast behaviors, anticipate needs, and hyper-personalize messages (Kaličanin et al., 2019). Machine learning offers great potential value for marketing-related applications (Villarroel & Silipo, 2021).

As Brei (2020) states, the application of cluster segmentation with the adoption of artificial intelligence technologies in marketing management contributes to decision-making related to customer and product management. In turn, Ullal et al. (2021) assert that the adoption of AI tools aims to provide companies with real-time information. Likewise, Jarek and Mazurek (2019) note that machine learning combines data science, quantitative methods, and computer programming.

AI contributes to the development of personalized marketing strategies, creating communication strategies tailored to customers based on their tastes and preferences, their geolocation data, and any other necessary information about customers. It also supports the segmentation of customer data, classifying groups of consumers by characteristics. Thus enabling companies to continuously evaluate over time, with a short-term perspective, the evolution of customers and to manage strategies dynamically along the way.

Additionally, AI allows companies to seize sales opportunities and, from the information provided by market segmentation, to observe closely the changes in consumer preferences. It is also possible to identify business opportunities based on the visits to a site of products or on purchases that consumers have made over time via digital media, which is known today as web analytics.

CONCLUSIONS

Segmentation is very useful in business marketing because it makes it possible to make decisions in real-time, prevent customer losses, and track trends in consumer tastes and preferences.

Using the data, it is also possible to observe how people develop feelings about a company's brands, whether these are positive or negative manifestations. COVID-19 health crisis has led to substantial changes in the behavior of consumers, affecting the way they buy and make purchasing decisions.

In the clustering paradigm that allows finding the Euclidean distance between groups of consumers in the field of business marketing, minimum distances indicate that there are no significant differences in consumer tastes and preferences; in contrast, large distances denote that there are significant differences.

Based on this information, four groups of consumers with markedly different perceptions regarding shopping were identified. One group stated that they saved money during sales periods, a second group indicated that they prefer to have a wide range of offers and prices to buy a product of their choice, a third group declared that when they go out shopping they take advantage of eating out with the family, while a fourth smaller group manifested that they were not interested in shopping because it is not convenient for their budget and they are not particularly motivated by it.

In conclusion, this study demonstrated the usefulness of artificial intelligence for estimating the optimal number of segments, supported by the Davies-Bouldin indicator based on performance and similarity measure of clusters, thus proving the research hypothesis.

ACKNOWLEDGMENT

To my *alma mater*, Universidad Nacional Jorge Basadre and its School of Commercial Engineering.

REFERENCES

- [1] Abakouy, R., En-Naimi, E. M., El Haddadi, A., & Lotfi, E. (2019). Data-driven marketing: How machine learning will improve decision-making for marketers. *SCA '19: Proceedings of the 4th International Conference on Smart City Applications*. <https://doi.org/10.1145/3368756.3369024>
- [2] Baltés, P. L. (2015). Content marketing - the fundamental tool of digital marketing. *Bulletin of the Transilvania University of Brasov*, 8(2), 111-118.
- [3] Bayoude, K., Ouassit, Y., Ardchir, S., & Azouazi, M. (2018). How machine learning potentials are transforming the practice of digital marketing: State of the art. *Periodicals of Engineering and Natural Sciences*, 6(2), 373-379. <https://doi.org/10.21533/pen.v6i2.526>
- [4] Brei, V. A. (2020). Machine learning in marketing: Overview, Learning Strategies, Applications, and Future Developments. *Foundations and Trends in Marketing*, 14(3), 173-236. <https://doi.org/10.1561/17000000065>
- [5] Casco, A. R. (2020). Efectos de la pandemia de COVID-19 en el comportamiento del consumidor. *Innovare: Revista de Ciencia y Tecnología*, 9(2), 98-105. <https://doi.org/10.5377/innovare.v9i2.10208>
- [6] Dzyabura, D., & Hauser, J. R. (2011). Active machine learning for consideration heuristics. *Marketing Science*, 30(5), 801-819. <https://doi.org/10.1287/mksc.1110.0660>
- [7] Eger, L., Komárková, L., Egerová, D., & Mičík, M. (2021). The effect of COVID-19 on consumer shopping behaviour: Generational cohort perspective. *Journal of Retailing and Consumer Services*, 61. <https://doi.org/10.1016/j.jretconser.2021.102542>
- [8] Ellickson, P. B., Kar, W., & Reeder, J. C. (2022). Estimating Marketing Component Effects: Double Machine Learning from Targeted Email Promotions. *Marketing Science*. <https://doi.org/10.1287/mksc.2022.1401>
- [9] Huang, M. H., & Rust, R. T. (2022). A Framework for Collaborative Artificial Intelligence in Marketing. *Journal of Retailing*, 98(2), 209-223. <https://doi.org/10.1016/j.jretai.2021.03.001>
- [10] Jain, S., & Kain, M. (2018). Prediction for Stock Marketing Using Machine Learning. *International Journal on Recent and Innovation Trends in Computing and Communication*, 6(4), 131-135.
- [11] Jarek, K., & Mazurek, G. (2019). Marketing and Artificial Intelligence. *Central European Business Review*, 8(2), 46-55. <https://doi.org/10.18267/j.cebr.213>
- [12] Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260. <https://doi.org/10.1126/science.aaa8415>
- [13] Kabir, M. Y., & Madria, S. (2021). EMOCOV: Machine learning for emotion detection, analysis and visualization using COVID-19 tweets. *Online Social Networks and Media*, 23. <https://doi.org/10.1016/j.osnem.2021.100135>
- [14] Kaličanin, K., Čolović, M., Njeguš, A., & Mitić, V. (2019). Benefits of Artificial Intelligence and Machine Learning in Marketing. *Sinteza 2019 - International Scientific Conference on Information Technology and Data Related Research*, 472-477. <https://doi.org/10.15308/sinteza-2019-472-477>
- [15] Kumar, A., Gawande, A., & Brar, V. (2020). Marketing Tactics in Times of Covid-19. *Vidyabharati International Interdisciplinary Research Journal*, 11(2), 263-266.
- [16] Mahajan, K. S., Jamsandekar, S. S., & Gurav, A. M. (2017). Machine Learning Approach for Marketing Intelligence: Managerial Application. *International Journal Of Engineering And Computer Science*, 6(7), 21929-21936.
- [17] Mahendiran, A., Saravanan, N., Subramanian, N., & Sairam, N. (2012). Implementation of K-means clustering in cloud computing environment. *Research Journal of Applied Sciences, Engineering and Technology*, 4(10), 1391-1394.

- [18] Mehta, S., Saxena, T., & Purohit, N. (2020). The New Consumer Behaviour Paradigm amid COVID-19: Permanent or Transient? *Journal of Health Management*, 22(2), 291-301. <https://doi.org/10.1177/0972063420940834>
- [19] Miklosik, A., Kuchta, M., Evans, N., & Zak, S. (2019). Towards the Adoption of Machine Learning-Based Analytical Tools in Digital Marketing. *IEEE Access*, 7, 85705-85718. <https://doi.org/10.1109/ACCESS.2019.2924425>
- [20] MR, A. (2021). Artificial Intelligence and Marketing. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(4), 1247-1256. <https://doi.org/10.17762/turcomat.v12i4.1184>
- [21] Rodríguez-Rodríguez, I., Rodríguez, J. V., Shirvanizadeh, N., Ortiz, A., & Pardo-Quiles, D. J. (2021). Applications of Artificial Intelligence, Machine Learning, Big Data and the Internet of Things to the COVID-19 Pandemic: A Scientometric Review Using Text Mining. *International Journal of Environmental Research and Public Health*, 18(16). <https://doi.org/10.3390/ijerph18168578>
- [22] Thiraviyam, T. (2018). Artificial Intelligence Marketing. *International Journal of Recent Research Aspects*, 19(4), 449-452.
- [23] Ullal, M. S., Hawaldar, I. T., Soni, R., & Nadeem, M. (2021). The Role of Machine Learning in Digital Marketing. *SAGE Open*, 11(4). <https://doi.org/10.1177/21582440211050394>
- [24] Villarroel Ordenes, F., & Silipo, R. (2021). Machine learning for marketing on the KNIME Hub: The development of a live repository for marketing applications. *Journal of Business Research*, 137, 393-410. <https://doi.org/10.1016/j.jbusres.2021.08.036>
- [25] Yıldırım, M., Geçer, E., & Akgül, Ö. (2021). The impacts of vulnerability, perceived risk, and fear on preventive behaviours against COVID-19. *Psychology, Health and Medicine*, 26(1), 35-43. <https://doi.org/10.1080/13548506.2020.1776891>