Circular Economy and Sustainability
https://doi.org/10.1007/s43615-023-00336-4

ORIGINAL PAPER

Examination of the Criticality of Customer Segmentation Using Unsupervised Learning Methods

Arpit Saxena[1] · Ashi Agarwal[2] · Binay Kumar Pandey[3] · Digvijay Pandey[4]

Received: 1 November 2023 / Accepted: 27 December 2023

Abstract

In the world, everything revolves around selling and buying to get something or to earn a living. Whoever is selling is a seller who needs a customer to sell the things. The customer went to a seller when the seller approached the customer. Long-term relationships with customers become more and more important as a marketing paradigm unfolds. To predict the customer–seller relationship or to analyze customer satisfaction, to efficiently identify and serve its customers depending on multiple variables, a corporation must segment its market because it has a finite number of resources. Clustering is a useful and popular method for market segmentation, which identifies the intended market and customer groupings, in the field of market research. This study demonstrates how to segment mall customers using machine learning methods. This is the unsupervised clustering problem, and three well-known algorithms—K-means, affinity propagation, and DBSCAN—will be discussed and contrasted. The primary goal of the study is to go through the fundamentals of clustering techniques while also touching on some more complicated ideas. The study also revealed that there are more female customers than male consumers, with women making up 56% of all customers. Males have a greater mean income than females ($62.2 k vs. $59.2 k). Additionally, male customers' median income ($62.5 k) is higher than female customers ($60 k). Both groups' standard deviations are comparable. With an annual income of roughly 140 k dollars, one male stands out in the group.

Keywords Customer segmentation · Cluster analysis · K-means · DBSCAN · EDA

✉ Digvijay Pandey
digit11011989@gmail.com

Arpit Saxena
arpitanasaxena@gmail.com

Ashi Agarwal
ashiagarwal31@gmail.com

Binay Kumar Pandey
binaydece@gmail.com

1 GNIOT Institute of Professional Studies, Greater Noida, Uttar Pradesh, India

2 ABES Engineering College, Ghaziabad, Uttar Pradesh, India

3 Department of Information Technology, College of Technology, Govind Ballabh Pant University of Agriculture and Technology, U S Nagar, Pantnagar, India

4 Department of Technical Education (Government of U.P), Kanpur, Uttar Pradesh, India

Published online: 09 January 2024

Introduction

Customer segmentation is the idea of dividing heterogeneity into homogeneous forms. In strategic marketing, segmentation has received a great deal of attention and is frequently employed. The vast majority of research in this field is concerned with the application or creation of various methodologies. In addition to the conventional multivariate statistical tools, researchers and practitioners are now able to apply very complex data analysis approaches thanks to the internet and database technologies, which have made a vast amount of data on markets and customers exploitable. Managers frequently rely on their gut instinct and conventional segmentation strategies based on socio-demographic characteristics, even though the application of segmentation is a topic of primary importance in marketing research and practice. As a tool that can direct businesses in the market toward more efficient ways to advertise their products and create new ones, customer segmentation has practically limitless possibilities. Customer segmentation is a subfield of market segmentation and analysis, market segmentation is the practice of dividing a market into groups of consumers with different needs, characteristics, or behaviors who may need different products or who may respond differently to various marketing methods. For segmentation to be effective, each segment must typically be evaluated for specific traits like stability, growth potential, size, accessibility, response, and whether the marketing initiatives targeted at that sector of the market are aligned with the goals and resources of the organization [1, 2]. Success depends on an organization's ability to create and maintain loyal and priceless client connections. Creating a comprehensive customer strategy is essential for maximizing customer value [3]. As one of the essential analytical techniques employed in data mining, clustering can be used independently to show data distribution, detect cluster features, and, if necessary, conduct further analysis on particular clusters [4–6].

Related Work

Cooil et al. [1] investigated a sizable data collection given by a global specialty goods retail business. To offer a significant alternate method of comprehending segment characteristics, the authors also provided an example of how principle components might be used as inactive covariates. Kim et al. [3] outlined a framework for calculating customer value and creating value-based client segments. An analysis of a cellular telecommunications company's case study will serve as an example of how to establish strategies for each customer segment after segmenting customers based on their value. Wu et al. [4] success depends on an organization's ability to create and maintain loyal and priceless client connections. Creating a comprehensive customer strategy is essential for maximizing customer value to examine various modes, researchers created a monetary matrix and a fluctuate-rate matrix. Marcus et al. [7] discussed and defined the importance of market and customer segmentation. The customer value matrix was used by the researcher to segment customers based on various factors to apply the RFM tool. Hamka et al. [8] illustrated how clusters can be tied to both demographic and psychographic groups. In addition to new mobile behavior theories that can be tested further, the data [9] offer fine-grained insights into market categories. Although the study is exploratory, it shows the value of consumer segmentation on smartphone measurement data. Hiziroglu et al. [10] investigated the most recent applications of soft computing techniques to segmentation issues that are founded on significant factors, such as those related to segmentation success. The critical review shows that the application of soft computing in segmentation problems is still in its infancy, and it is probable that the knowledge that can be generated by these studies will not be sufficient. It was suggested that additional research be conducted to create more results that are acceptable to managers and can be more readily interpreted by managers based on these findings. Hosseini et al. [11] K-means clustering and the RFM model were used to classify customers according to their value. The assessment of modifications throughout various periods is then completed. The novelty of this research is in its improvement of the accuracy of forecasts based on prior customer behavior by taking into account time and trend of customer value changes. The POS customer transactions were employed by the researchers for this aim. Bayer et al. [12] thanks to modern segmentation techniques, which enable precise targeting with information on each customer's retention and value drivers, it is feasible for each customer to be a part of a micro-segment. The study's secondary purpose of increasing retention and growth is to improve company planning, where each area may be given particular growth and retention goals. Customer value segmentation, customer behavior segmentation, customer life cycle segmentation, and customer migration segmentation were the four techniques used to divide up the customer base. Das et al. [13] used data mining techniques [14] to thoroughly investigate client segmentation. It is an organized study of segmentation methods that use supervised, unsupervised, and other data mining approaches.

Data Analysis Outline

The process of classifying [15] customers is known as customer segmentation according to shared traits so that businesses can more successfully market to each group. To optimize [16] each customer's value to the company, it is important to segment customers to determine how to interact with them. Customer segmentation may enable businesses to reach out to each customer in the most efficient way. With the use of segmentation [17], a business may develop and communicate [18] customized marketing strategies that will appeal to particular client segments, enhancing customer relations and customer service [19]. Now, the main question is how this segmentation can be carried out. The answer is in favor of market and customer base; to segment its customer base, a business must collect detailed information about its clients [20], or data, and then analyze that information [21] to find trends. Data from purchases may reveal information on things like job title, location, and products bought, among other things. Whenever a customer requested a membership for a mall patron. They must complete a membership form, from which we may obtain information such as Name, Age, Gender, and Annual Income [22]. I then compute a spending score for each customer, ranging from 1 to 100, based on how frequently they visit the business and how much money they spend there. Wang et al. [23] employed a self-organizing map with an improved social spider optimization technique [24], an unsupervised deep learning model [25], to segment customers effectively. Modified social spider optimization, a swarm intelligence model [26], is used in a feature engineering process to analyze the customer data and pick the behavioral attributes of the consumer.

Properties of Cluster

As an easy illustration, a bank wants to provide its clients with credit cards. Currently, they examine each customer's details to determine which offer should be extended to a particular customer. Millions of people could potentially be the bank's clients now. Does it make sense to consider each customer's specifics independently before making a choice? Not! It will take a long time because it is a manual process.

How then can the bank respond? One possibility is to divide its clients into various categories. For instance, the bank can classify the clients according to their level of income [27]. For this segmentation, some properties should be taken into account as follows:

1. A cluster's data points ought to share similarities. Using the aforementioned example, let me explain: It stands to reason that if the customers in a given cluster are not all alike, then their needs can differ. They might not like the same offer if the bank makes it, and their interest in the bank might decline. Not optimal. The bank can employ targeted marketing more effectively when related data points are contained in the same cluster.
2. The ideal situation is for the data points from different clusters to be as different as possible.

Experiment Analysis and Modelling

Customer clustering based on shared traits is an unsupervised learning process [28] because there is no goal variable for each observation. The dataset Mall has the architecture like, there are 200 observations overall, and there are 5 variables for each observation. The dataset has columns for Customer ID, Gender, Age, Annual Income, and Spending Score. No values are missing (good day for us XD) and one categorical variable—Gender—is present (Fig. 1).

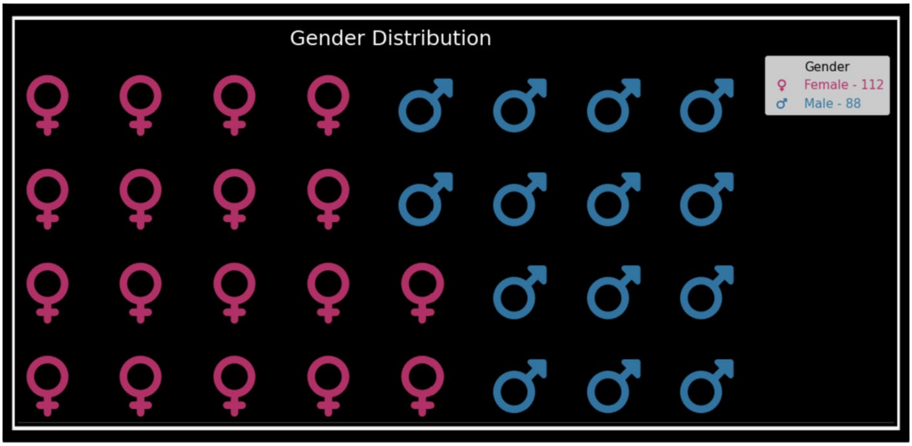


Fig. 1 Gender distribution

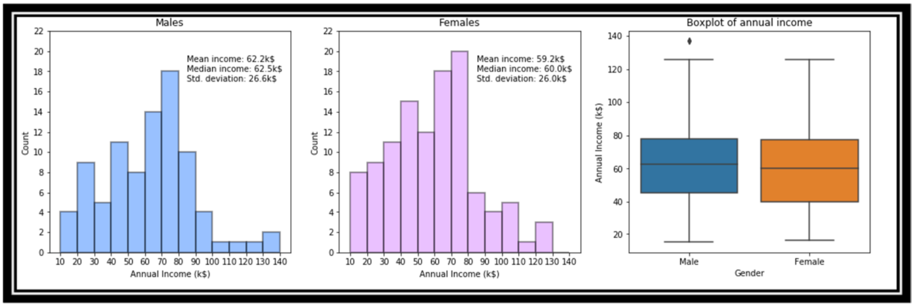


Fig. 2 Annual income gender wise

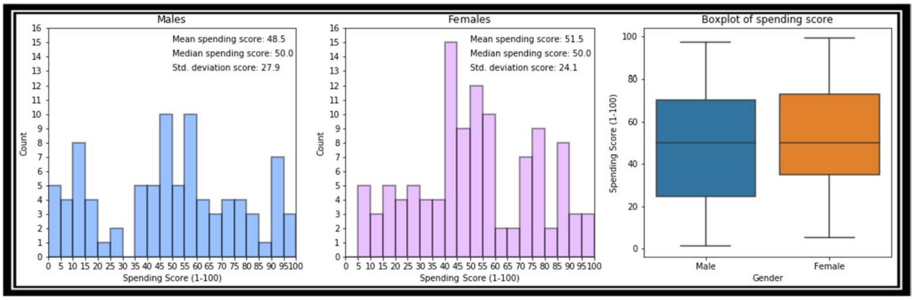


Fig. 3 Annual spending score gender wise

Exploratory Data Analysis

A fundamental statistical analysis [29] of a particular database is presented in this section. It is an essential part of every analysis since it enables a better comprehension of the underlying information. The distribution and correlation sections make up the bulk of this section. Here, categorical features [30] are defined; if the attribute has less than six unique elements, it is a numerical feature. A typical approach for this division of features can also be based on the datatypes of the elements of the respective attribute [32-48].

E.g., datatype = integer, attribute = numerical feature; datatype = string, attribute = categorical feature.

For this dataset, as the number of features is less, we can check the dataset manually as well.

The dataset shows that there are more female customers than male consumers, with women making up 56% of all customers (Fig. 2).

Males have a greater mean income than females ($62.2 k vs. $59.2 k). Additionally, male customers' median income ($62.5 k) is higher than female customers' income ($60 k). Both groups' standard deviations are comparable. With an annual income of roughly 140 k dollars, one male stands out in the group. The K-S test reveals no statistically significant difference between the two groups (Fig. 3).

Women have a higher mean spending score (51.5) than men do (48.5). There is no evidence to support the null hypothesis, according to the K-S test *p*-value, but the

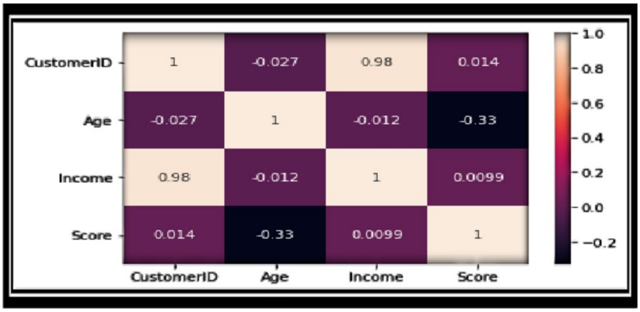


Fig. 4 Correlation matrix

evidence is not as strong as in earlier comparisons. I will then determine the median salary for each age group (Fig. 4).

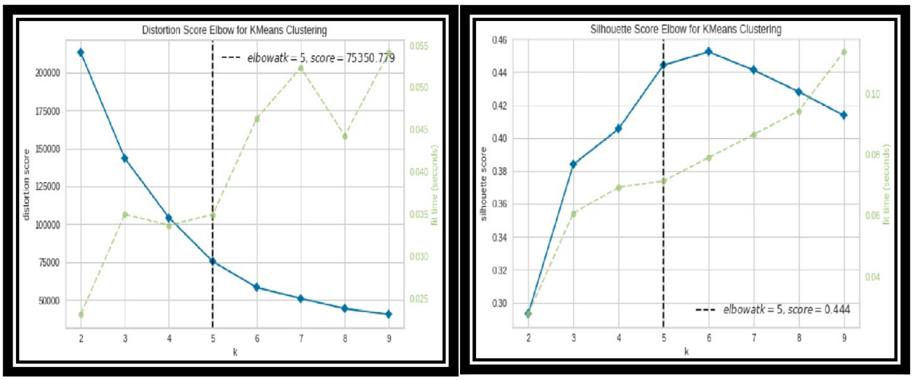
K-Means Clustering

In K-means, the Euclidean distance between an object [31] and the cluster centroid, also known as the cluster center, is used to group things into clusters [19]. However, neither the number of clusters nor the nature of the clusters is known in advance. We therefore proceed in two stages:

i. Finding the ideal number of clusters is the first step.
ii. For every cluster, we choose the beginning values.

In the study, researchers have used the elbow method to find the optimal number of clusters as shown in Fig. 1.

As the number of clusters rises, the distortion score decreases, as seen in the graph above. However, it is impossible to see an obvious "elbow." Five clusters are proposed by



(a) (b)

Fig. 5 **A** Distortion score by elbow method, **b** distortion score by silhouette method the underlying algorithm. It seems fair to offer a choice of five or six clusters. The silhouette score method indicates the best options would be 5 or 6 clusters (Fig. 5).

The squared Euclidean distance is a popular measure for grouping samples with continuous variables. Similarity is the reverse of distance. In m-dimensional space, the distance between two points x and y is determined by:

$$d(x,y)^2 - \sum_{j=1}^{m} (x_j - y_j)^2 = \|x - y\|_2^2$$

The K-means algorithm, which employs an iterative process to reduce the intra-cluster sum of squared errors (SSE), also known as cluster inertia, can be thought of as a straightforward optimization problem based on this Euclidean distance metric.

$$SSE = \sum_{i=1}^{n} \sum_{j=1}^{k} w^{(i,j)} \left| \left| x^{(i)} - \mu^{(j)} \right| \right|_2^2$$

The clusters of customers in the mall dataset are as follows based on their annual income and spending scores (Fig. 6):

We gain a very clear understanding of the various customer segments in the mall thanks to our clustering analysis. The best variables or attributes to identify the client segments at a mall are believed to be their annual income and spending score, which is used to classify customers into five separate clusters. Customers in cluster 0 fall into the low yearly income (40 k) and consequently low spending score (40) categories. Customers in cluster 1 are those with high yearly incomes (> \$50,000) and high spending scores (> \$70). We can identify the group of persons with an average annual income (40 k to 70 k) and an average

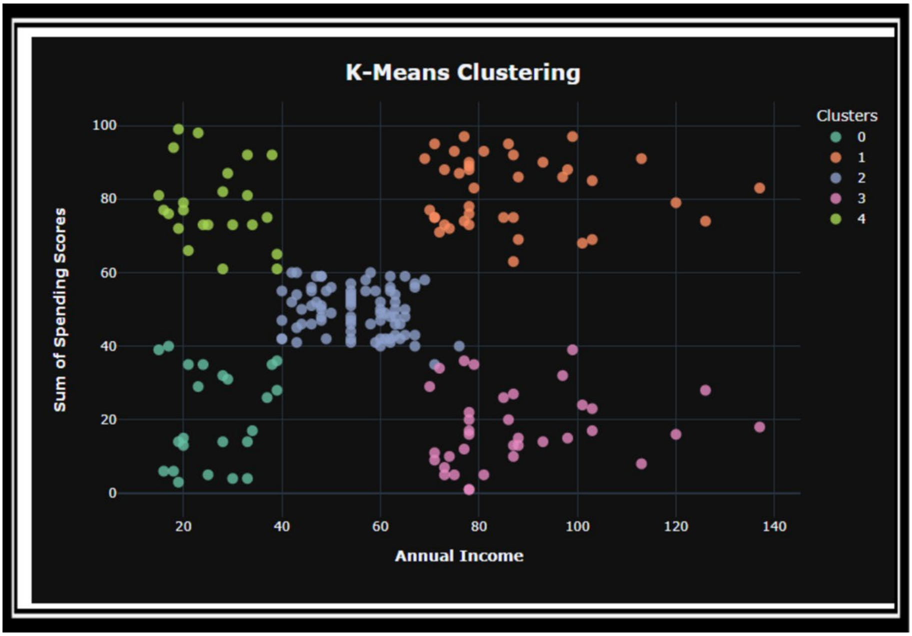


Fig. 6 Clustering by K-means

spending score using cluster 2 (40–60). The customers in cluster 4 have a high spending score despite having a modest annual income. Despite having a high annual income, those in the final group, cluster 5, have poor spending scores.

Hierarchical Clustering

Hierarchical clustering, also known as hierarchical cluster analysis, is the act of grouping objects into groups based on how similar they are to one another (Fig. 7). The outcome is a collection of clusters, each of which stands out from the others while exhibiting the majority of the same traits.

The advantages of hierarchical clustering are that there is no presumption of a specific number of clusters, unlike K-means and possibly match meaningful taxonomy. While the disadvantages of hierarchical clustering state that combining two clusters is a decision that cannot be changed after it has been made and O(n2log(n)) is incredibly slow for huge data collections.

Hierarchical clustering working:

i. Cluster each data point.
ii. Combine the two groups that are closest to one another.
iii. Continue with step 2 up until there is just one cluster.

The dendrogram up top shows Ward's linkage hierarchical clustering, where the distance between clusters is the product of the squared differences between each cluster (Fig. 8). We must draw a horizontal line across the dendrogram to get the total number of clusters. The total number of clusters is determined by counting the number of vertical lines that this horizontal line (yellow) cuts. The ideal number of clusters is five because a horizontal line cuts five vertical lines.

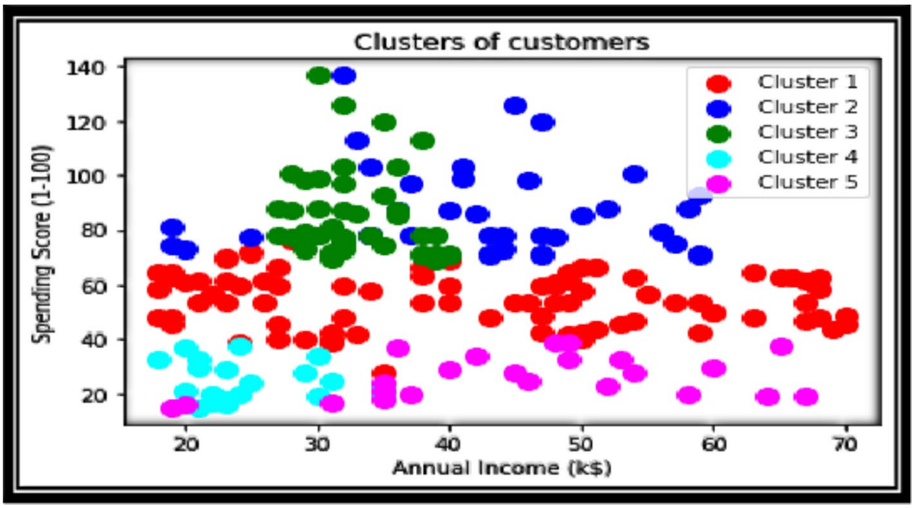


Fig. 7 Clustering by hierarchical method

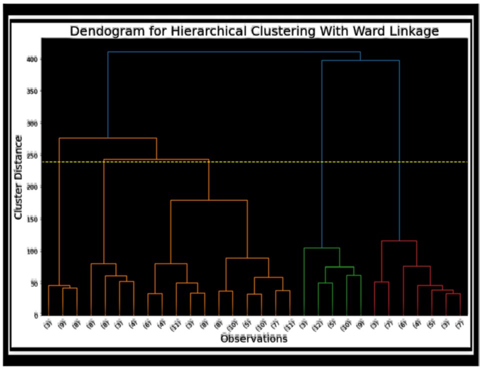


Fig. 8 Dendogram for hierarchical clustering

Agglomerative Clustering

This method is described as "bottom-up" since each observation begins in its cluster and pairs of clusters are connected as one travel up the hierarchy (Fig. 9).

Affinity Propagation

By transferring messages [32] between data points until convergence, affinity propagation builds clusters (Fig. 10). The number of clusters is not a requirement for affinity propagation, unlike other conventional clustering techniques. To put it simply, each data point in affinity propagation sends signals to every other data point telling its recipients of the relative

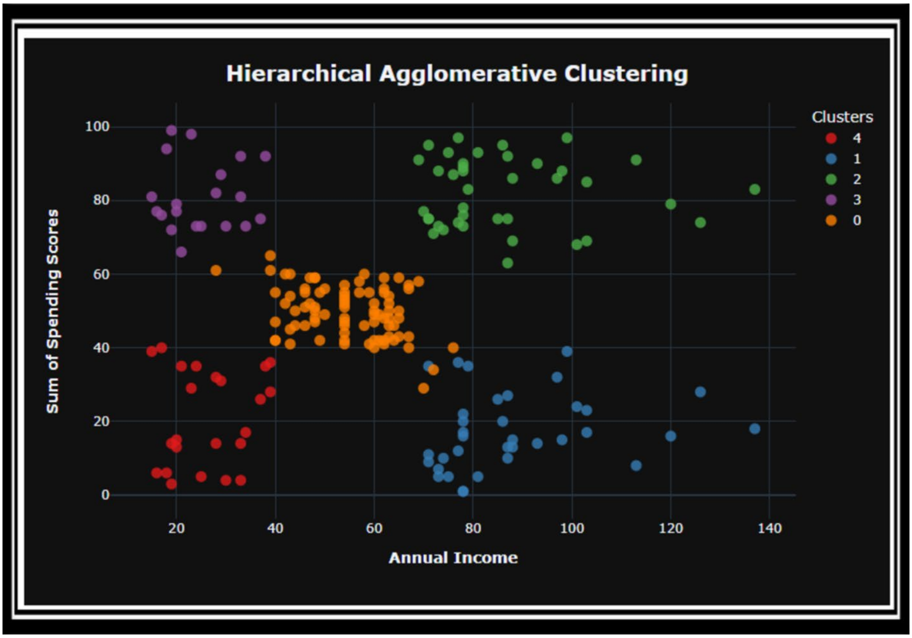


Fig. 9 Clustering by agglomerative method

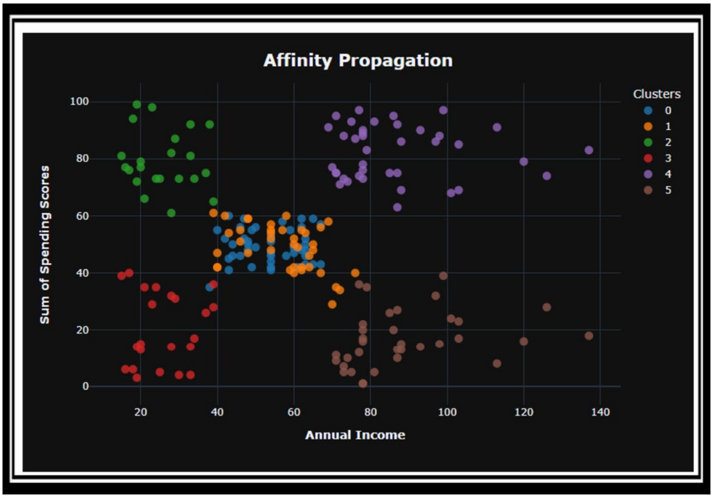


Fig. 10 Clustering by affinity propagation

attraction of each recipient to the sender [33]. Then, given the appeal of the communications, it has received from all other senders, each target replies to each sender alerting them of its willingness to engage with them. Using the availability messages, it has received from all of the targets, senders respond to the targets with messages alerting each target of its updated relative attractiveness to the sender. Messages are passed back and forth until an agreement is achieved. One of its targets becomes the point's example once the sender is connected to that target. The same cluster contains all the points that share the same exemplar.

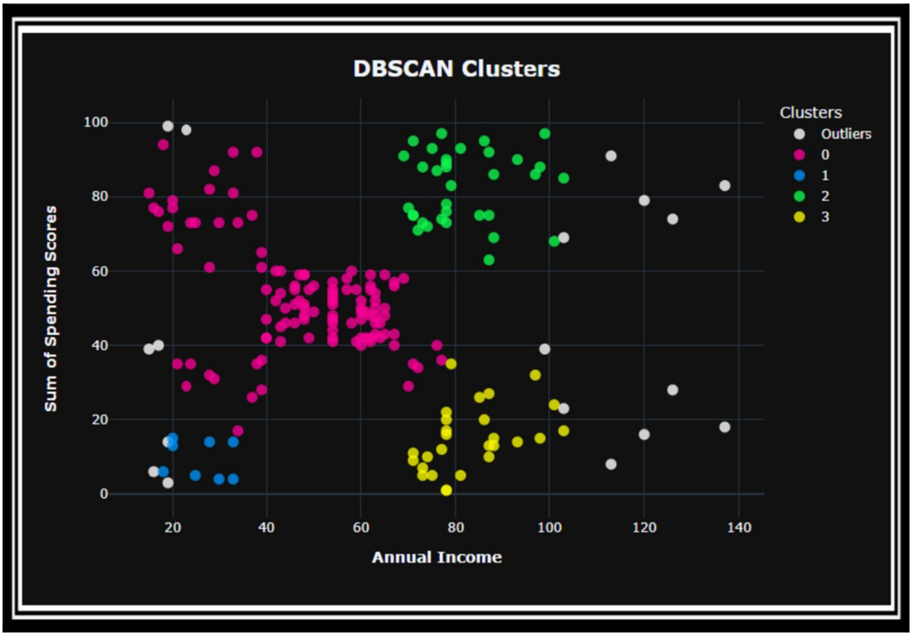


Fig. 11 Clustering by DBSCAN

DBSCAN

The unsupervised learning [34] method known as DBSCAN is frequently used in model construction and machine learning algorithms [35] (density-based spatial clustering of applications with noise). In machine learning [36], it is a method of clustering that helps to distinguish between clusters with high and low densities. DBSCAN does a great job at distinguishing between areas of the data that have a high density of observations and those that do not as a density-based clustering algorithm [20].

DBSCAN's functionality is to divide the datatest n- dimensionally.

Each data point in the dataset is surrounded by an n-dimensional shape created by DBSCAN, which then counts how many data points fit inside the shape and gives the findings. DBSCAN identifies this shape as a cluster. DBSCAN travels each point in the cluster, counting the number of neighboring data points as it goes, increasing the cluster repeatedly (Fig. 11).

Conclusion

Customer segmentation is the practice of classifying customers based on characteristics they have in common so that organizations may market effectively to each group. By segmenting the audiences, marketers can more effectively target various audience subgroups with their promotional activities. The activities might include both communications and product development. Customer segmentation helps it in creating and distributing customized marketing materials that will appeal to customer segments and developing stronger connections with your customers. Examining different price points and concentrating on the most lucrative customers, enhancing client interactions to build customer segments, a business must collect precise data about its consumers in all fields like telecommunication, mobile, and general [20] and analyze it to look for trends. When determining how to sell to distinct customer segments and what goods or services to promote to them, a corporation can use common traits discovered from various consumer segments as a guide. Customer segmentation can be used by any business, regardless of their size, sector of operation, or method of sales (online or offline). It begins with gathering and analyzing data and ends with taking appropriate and effective action in response to the knowledge gained. This dataset is great for practicing with difficulties in unsupervised learning [23, 37]. It provides the opportunity to work on a business issue that might be used to create several approaches for increasing sales. Finding the hidden patterns in the data using EDA is particularly helpful for unsupervised learning challenges. As the number of features and size of the data grows, the complexity of the visualizations might be a challenge.

The selection of the hyperparameter k for the K-means clustering algorithm is essential, and it is accomplished with the aid of statistical testing. When trained on the original dataset or, in this case, the normalized dataset, the model performance does not significantly differ. In the future, this study can be further applied by different researchers by using different techniques other than unsupervised learning methods. The other techniques can be soft computing and deep learning. However, some data is missing in the database because of the lack of tools and techniques to extract customer data [23].

Acknowledgements The authors thank ABES Engineering College, Ghaziabad, GIPS GREATER NOIDA, DTE, and GBPUAT for providing facilities and support to complete the project.

Author Contribution All authors are equally contributed in manuscript. All authors reviewed the manuscript.

Data Availability The raw or processed data required to reproduce these findings cannot be shared at this time as the data also forms part of ongoing work.

Declarations

Conflict of Interest The authors declare no competing interests.

References

1. Cooil B, Aksoy L, Keiningham TL (2008) Approaches to customer segmentation. J Relationship Marketing 6(3–4):9–39. https://doi.org/10.1300/j366v06n03_02
2. Balmeet Kaur, Pankaj Kumar Sharma, (2019) Implementation of customer segmentation using integrated approach", International Journal of Innovative Technology and Exploring Engineering (IJITEE), Blue Eyes Intelligence Engineering & Sciences Publication. 8(6S):770- 772, ISSN 2278–3075
3. Kim S, Jung T, Suh E, Hwang H (2006) Customer segmentation and strategy development based on customer lifetime value: a case study. Expert Syst Appl 31(1):101–107. https://doi.org/10.1016/j.eswa.2005.09.004
4. Wu J, Lin Z (2005) Research on customer segmentation model by clustering. Proceedings of the 7th International Conference on Electronic Commerce - ICEC '05. https://doi.org/10.1145/1089551.1089610
5. Ziafat H, Shakeri M (2014) Using data mining techniques in customer segmentation. J Eng Res App 4(9):70–79
6. Hosseini M, Shabani M (2015) New approach to customer segmentation based on changes in customer value. J Marketing Anal 3:110–121
7. Marcus C (1998) A practical yet meaningful approach to customer segmentation. J Consum Mark 15(5):494–504. https://doi.org/10.1108/07363769810235974
8. Hamka F, Bouwman H, De Reuver M, Kroesen M (2014) Mobile customer segmentation based on smartphone measurement. Telematics Inform 31(2):220–227. https://doi.org/10.1016/j.tele.2013.08.006
9. Pandey BK, Pandey D, Wairya S, Agarwal G (2021) An advanced morphological component analysis, steganography, and deep learning-based system to transmit secure textual data. Int J Distributed Artificial Intell (IJDAI) 13(2):40–62
10. Hiziroglu A (2013) Soft computing applications in customer segmentation: state-of-art review and critique. Expert Syst Appl 40(16):6491–6507. https://doi.org/10.1016/j.eswa.2013.05.052
11. Hosseini M, Shabani M (2015) New approach to customer segmentation based on changes in customer value. J Marketing Anal 3(3):110–121. https://doi.org/10.1057/jma.2015.10
12. Bayer J (2010) Customer segmentation in the telecommunications industry. J Database Mark Cust Strategy Manag 17(3–4):247–256. https://doi.org/10.1057/dbm.2010.21
13. Das S, Nayak J (2022) Customer segmentation via data mining techniques: state-of-the-art review. Computational Intelligence in Data Mining, 489–507. https://doi.org/10.1007/978-981-16-9447-9_38
14. Babu SZD, Pandey D, Naidu GT, Sumathi S, Gupta A, Bader Alazzam M, Pandey BK (2022) Analysation of big data in smart healthcare. In Artificial intelligence on medical data: proceedings of international symposium, ISCMM 2021 (243–251). Singapore: Springer Nature Singapore.
15. Revathi TK, Sathiyabhama B, Sankar S, Pandey D, Pandey BK, Dadeech P (2022) An intelligent model for coronary heart disease diagnosis. Networking technologies in smart healthcare: innovations and analytical approaches, 234.
16. Boopathi S, Pandey BK, Pandey D (2023) Advances in artificial intelligence for image processing: techniques, applications, and optimization. In Handbook of research on thrust technologies' effect on image processing (73–95). IGI Global.
17. Pandey BK, Pandey D, Alkhafaji MA, Güneşer MT, Şeker C (2023) A reliable transmission and extraction of textual information using keyless encryption, steganography, and deep algorithm with cuckoo optimization. In Micro-electronics and telecommunication engineering: proceedings of 6th ICMETE 2022 (629–636). Singapore: Springer Nature Singapore.

18. Kirubasri G, Sankar S, Pandey D, Pandey BK, Nassa VK, Dadheech P (2022) Software-defined networking-based Ad hoc networks routing protocols. Software defined networking for Ad Hoc networks. Springer International Publishing, Cham, pp 95–123
19. Dinesh Kumar Yadav, Rati Shukla, Vikash Yadav, (2021) An efficient collaborative recommender system for textbooks using silhouette index and K-means clustering technique. Int J Adv Intell Paradigms, InderScience Publication. 19(2):233- 242, ISSN 1755–0394 https://doi.org/10.1504/IJAIP.2021.115251
20. Vikash Yadav, Rati Shukla, Aprna Tripathi, Anamika Maurya. (2021) A new approach for movie recommender system using K-means clustering and PCA. J Scientific Industrial Res, NISCAIR-CSIR, India. 80(2):159- 165, ISSN 0975–1084
21. Pandey D, Pandey BK, Wairya S (2021) Hybrid deep neural network with adaptive galactic swarm optimization for text extraction from scene images. Soft Comput 25:1563–1580
22. Ahmad AYB, Gongada TN, Shrivastava G, Gabbi RS, Islam S, Nagaraju K (2023) E-commerce trend analysis and management for industry 5.0 using user data analysis. Int J Intell Systems App Eng 11(11s):135–150
23. Wang Chenguang (2022) Efficient customer segmentation in digital marketing using deep learning with swarm intelligence approach. Inform Process Management 59(6):103085. https://doi.org/10.1016/j.ipm.2022.103085
24. Pandey D, Pandey BK (2022) An efficient deep neural network with adaptive galactic swarm optimization for complex image text extraction. In Process mining techniques for pattern recognition (121–137). CRC Press.
25. Pandey BK, Pandey D, Wariya S, Agarwal G (2021) A deep neural network-based approach for extracting textual images from deteriorate images. EAI Endorsed Transact Industrial Networks Intelligent Systems 8(28):e3–e3
26. Kumar Pandey B, Pandey D, Nassa VK, Ahmad T, Singh C, George AS, Wakchaure MA (2021) Encryption and steganography-based text extraction in IoT using the EWCTS optimizer. The Imaging Science Journal 69(1–4):38–56
27. Ahmad AYB (2019) Empirical analysis on accounting information system usage in banking sector in Jordan. Acad Acc Finan Stud J 23(5):1–9
28. Anand R, Khan B, Nassa VK, Pandey D, Dhabliya D, Pandey BK, Dadheech P (2023) Hybrid convolutional neural network (CNN) for Kennedy Space Center hyperspectral image. Aerospace Systems 6(1):71–78
29. Pandey BK, Pandey D, Gupta A, Nassa VK, Dadheech P, George AS (2023) Secret data transmission using advanced morphological component analysis and steganography. Role of data-intensive distributed computing systems in designing data solutions. Springer International Publishing, Cham, pp 21–44
30. Khan B, Hasan A, Pandey D, Ventayen RJM, Pandey BK, Gowwrii G (2021) 12 Fusion of datamining and artificial intelligence. Machine learning and iot for intelligent systems and smart applications, 201.
31. Pandey D, Wairya S, Sharma M, Gupta AK, Kakkar R, Pandey BK (2022) An approach for object tracking, categorization, and autopilot guidance for passive homing missiles. Aerospace Systems 5(4):553–566
32. Pandey JK, Jain R, Dilip R, Kumbhkar M, Jaiswal S, Pandey BK, ... Pandey D (2022) Investigating role of IoT in the development of smart application for security enhancement. In IoT based smart applications (219–243). Cham: Springer International Publishing.
33. Pramanik S, Pandey D, Joardar S, Niranjanamurthy M, Pandey BK, Kaur J (2023) An overview of IoT privacy and security in smart cities. In AIP Conference Proceedings (2495:1). AIP Publishing.
34. Pandey BK, Pandey D (2023) Parametric optimization and prediction of enhanced thermoelectric performance in co-doped CaMnO3 using response surface methodology and neural network. J Mater Sci: Mater Electron 34(21):1589
35. Meslie Y, Enbeyle W, Pandey BK, Pramanik S, Pandey D, Dadeech P, ... Saini A (2021) Machine intelligence-based trend analysis of COVID-19 for total daily confirmed cases in Asia and Africa. In Methodologies and applications of computational statistics for machine intelligence (164–185). IGI Global.
36. Singh H, Pandey BK, George S, Pandey D, Anand R, Sindhwani N, Dadheech P (2022) Effective overview of different ML models used for prediction of COVID-19 patients. In Artificial intelligence on medical data: proceedings of international symposium, ISCMM 2021 (185–192). Singapore: Springer Nature Singapore.
37. Abdulhafedh A (2021). Incorporating k-means, hierarchical clustering and pca in customer segmentation. J City Develop. 3(1):12–30 Wang, C. (2022).

38. Yash Veer Singh, Pushpendra Singh, Shadab Khan, Ram Sewak Singh, (2022) "A machine learning model for early prediction and detection of sepsis in intensive care units patients. J Healthcare Eng , Hindawi, 2022(Special Issue):1- 11, ISSN: 2040–2295 (Print) ISSN: 2040–2309 (Online) https://doi.org/10.1155/7158 https://doi.org/10.1155/2022/9263391
39. Harivans Pratap Singh,Priti Dimri,Shailesh Tiwari, Manish Saraswat, (2020) Segmentation techniques through machine based learning for latent fingerprint indexing and identification. J Scientific Industrial Res, NISCAIR-CSIR, India. 79(03):201- 208, ISSN 0975–1084 JSIR Vol.79(03) [March 2020]
40. Harshit Mishra, Suraj Kumar Pathak, Aastik Srivastava, Manish Srivastava, (2020) Handwritten digits recognition using machine learning algorithms. Int Res J Eng Technol (IRJET), IRJET. 7(5):8073-8079, ISSN 2395–0056
41. Dullaghan C, Rozaki E (2017) Integration of machine learning techniques to evaluate dynamic customer segmentation analysis for mobile customers. arXiv preprint arXiv:1702.02215.
42. Sari JN, Nugroho LE, Ferdiana R, Santosa PI (2016) Review on customer segmentation technique on ecommerce. Adv Sci Lett 22(10):3018–3022
43. Agarwal A (2022) CNN and PCA in image fusion: a comparative statistical analysis. Int J Performability Eng 18(12):885
44. Agarwal A, Pandey BK, Devi P, Kumar S, Paramashivan MA, Agarwal R, Dadheech P (2023) Comparison of the theoretical and statistical effects of the PCA and CNN image fusion approaches. In: Handbook of research on thrust technologies' effect on image processing. IGI Global, pp 193–205
45. Singh SJ, Sinha A, Yadav S (2022) Performance analysis of machine learning algorithms for erythemato-squamous diseases classification. In: 2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE), Ballari, India, pp 1–6. https://doi.org/10.1109/ICDCECE53908.2022.9793000
46. Singh G, Misra DK, Tanwar C, Narang P (2022) Visualization and analysis of populations. In: 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON), Faridabad, India, pp 364–368. https://doi.org/10.1109/COM-IT-CON54601.2022.9850886
47. Kumar S, Kumar R, Haider M, Dubey A (2021) A comparative study of classifier algorithms for Twitter's sentiment based spam detection. In: 1st International Conference on Computational Research and Data Analytics (ICCRDA 2020), India, Vol. 1022(1), pp 12016–12024. https://doi.org/10.1088/1757-899X/1022/1/012016
48. Devasenapathy D, Madhumathy P, Umamaheshwari R, Pandey BK, Pandey D (2024) Transmission-efficient grid-based synchronized model for routing in wireless sensor networks using Bayesian compressive sensing. SN Comput Sci 5(1):1–11