*i*Journals: International Journal of Software & Hardware Research in Engineering (IJSHRE)
ISSN-2347-4890
Volume 10 Issue 6 June 2022

Scrutiny of The Results Obtained from The Unsupervised Machine Learning Algorithms for The Customer Segmentation Problem

Authors: Ariha Mehta(1), Reetu Jain(2)

(1)Class of 12th, Jamnabai Narsee School, N.S. Road No. 7, JVPD Scheme, Vile Parle West. Mumbai - 400049, India.
(2)Mentor, On My Own Technology Pvt. Ltd., 2A, Pace House, 7 Swastik Society, Gulmohar Road, Vile Parle West, Mumbai - 400049, India.
Email-id: reetu.jain@onmyowntechnology.com

Abstract: Customer segmentation (CS) is a strategy to focus on the group of similar spending behavior customers. The advantage of CS is to closely align their policy and tactics to better target the customers. Segmenting customers on the basis of their spending behavior can be done with the help of machine learning as it is a great tool for analyzing and finding patterns in the dataset. The study is conducted in two phases. The first phase includes an exploratory data analysis (EDA) that helps in understanding the customer's traits and their spending behavior. The second phase of the study includes the development of the ML model. For the problem of CS, there are no label outputs for which the data are trained. Hence unsupervised learning (UL) is the most preferred ML tool for the CS problem. Groupings of the customers are computed using two UL models namely k-means and density-based spatial clustering of applications with noise (DBSCAN). The former model grouped the data into 5 clusters whereas the later model grouped it into 3 clusters along with some data identified as noise. The noise from the context of the problem of CS refers to those customers for who cannot be placed in any clusters. The application of UL in this study may further open up the potential for other applications in the same industry.

Keywords: k-means clustering, Density-based spatial clustering of applications with noise (DBSCAN), Unsupervised machine learning, Customer segmentation (CS), Performance visualizations

1. Introduction

Customer segmentation (CS) is the process of dividing customers into groups based on common characteristics so companies can market to each group effectively. It is an effective tool for businesses to closely align their strategy and tactics with, and better target their customers. Malls or shopping complexes are often indulged in the race to increase their customers and hence making huge profits. The primary strategy followed by the manager of the malls is to devise a common policy to the customers that showed similar spending behavior. However in order to divide the customers into groups of similar spending behavior, the application of machine learning (ML) comes in handy.

ML techniques are great tools for analyzing customer data and finding insights and patterns which can lead to rational conclusions and can help in the process of decision-making. However, for the problem of CS, there are no label outputs for which the data are trained. Hence unsupervised learning (UL) is the most preferred ML tool for the CS problem in a mall.

1.1. Literature review

This section of the paper summarizes the contemporary work done by the researchers in the field of machine learning (ML) applied to the customer segmentation problem (CS). The CS problem is a case of unsupervised learning (UL) where the customers are grouped together by a common pattern. Kansal et. al. (2018) applied 3 different clustering algorithms on 200 samples from a shop. The results of clusters obtained from k-means and agglomerative methods were the same while the mean shift method showed two new clusters: High buyers and frequent visitors and high buyers and occasional visitors [1]. By using K-means clustering it is concluded that customers who have high-income levels and spending scores, are suitable targets for implementing market strategies [2]. By using a two-step cluster analysis and subsequent MANOVA during a recessionary period it was found that 'young enthusiasts' are an attractive target group while 'adverse reluctant customers' are a challenge [3]. Findings show that criteria having the highest impact on the preferences of young people are mall campaigns for loyal customers, traffic in its locality, its parking facilities, facilities for disabled people, quality of its locality and quality of people visiting the mall [4]. This study suggests that an adequate tenant mix and a pleasant, appealing environment greatly attract customers, but the convenience of the shopping mall and the communication activities are not as significant [5]. A structured questionnaire captured factors like Lucky and Bumper offers, Frequent and Warranty Offers, Monetary and Quantity Benefit Offers, Gift and Exchange offers and Discount offers. It was found that students, between the ages of 20-30 yrs, who were dependent on their parents were conscious about sales promotion in shopping malls [6]. Findings show a significant transitional difference between three intra-cohort groups: adolescents, emerging adults and young adults. Emerging adults may be lost unless retailers are able to engage them [7]. Research from a mall in South Africa shows gender differences; the time and money spent by women in malls is comparatively higher than men [8]. This research paper identified 3 main shopping motives: hedonic, efficiency, and accomplishment and 3 shopper segments: hedonists, achievers, and efficient shoppers. It showed the significant differences among them [9]. A quantitative research uncovered that the satisfaction level among customers tends to be high. It also helped to understand the relation of age and gender on customers' satisfaction and loyalty in the mall [10]. Using a consumer household survey and retail audit, a study shows that consumers regard car convenience as a significant determinant of where to shop [11]. A research study has revealed the various reasons why customers visit a mall and with what frequency. Customers view malls not only as shopping centers but also as an entertainment source [12].A study's findings show that physical environment and personal interaction encounters play an important role in customers' omnichannel experience in smart malls. Using a customer-centric omnichannel strategy can help gain a better understanding of the customers [13].A research in a mall in South Africa found six segments associated with shopping preferences: 'Low Engagement Shoppers', 'Mean Oriented Shoppers', 'Brand Oriented Shoppers', 'Utilitarian Oriented Shoppers' and 'Ascetic Oriented Shoppers'[14]. This paper shows the impact of socioeconomic profile with variables like shop image, customer satisfaction w.r.t price, quality, payment mode using exploratory factor analysis [15]. Findings of a study reveal four dimensions: mall brand mix, consumer behavior, company attributes and market structure; and 21 critical success factors which help in forming the perfect strategy [16]. A paper aiming to investigate materialism as a retail shopping motive, through a mall intercept survey, divides Indian shoppers into four clusters: balanced shoppers, materialist shoppers, hedonistic shoppers and value shoppers [17]. A research in India, using hierarchical and K-means clustering methods, found three cluster groups: hygienic, extended and prudent all of which showed significant differences [18]. In tier II cities, Findings have identified six small dimensions -mall environment, convenience, mall staff, mall hygiene, entertainment and security- that help predict customer satisfaction [19].A research done on 400 Indian shoppers shows the significant difference between different age groups in terms of shopping behavior. But men and women behave similarly in terms of visit frequency and time and money spent per visit [20]. A research in a mall in Ahmedabad exhibits how visual merchandising (window display, in-store form/mannequin display, floor merchandising and promotional signage) results in impulse purchases [21]. This paper suggests that the atmosphere of the mall, its image, events organized, discounts offered greatly attract customers [22,25]. This study states that younger generations are more attracted to mall events and promotional schemes than older people [23]. Findings from this paper depict the relationship between income levels of young shoppers and its relation to impulsive buying. Antecedent factors have formed 5 shopper segments [24]. More research papers are reviewed for the study but limiting the literature section to the most recent and relevant papers.

1.2. Motivation and novelty

From the literature reviewed for the study some of the gaps that are identified are as follows:

1. Although there is much literature that involves the application of ML in the problem of CS, yet there is very little research done in the field of CS in malls.
2. The literature reviewed for the study mostly applies centroid based clustering algorithms without mentioning its upper hand in that over other clustering algorithms.
3. Above that, the existing literature either performs exploratory data analysis (EDA) to identify the spending behavior of the customers or groups them into clusters depending on a certain pattern. There is very little research paper that simultaneously performs the two tasks.

In the present study, the first gap identified in the literature will be addressed by collecting the dataset from the customers of the malls. The second gap identified in the literature will be addressed by solving the CS problem by k-means and DBSCAN algorithm. The k-means algorithm is a centroid based clustering algorithm and the DBSCAN is a density-based clustering algorithm. The motive behind solving the same problem by two algorithms is to know the advantage of centroid based clustering algorithms over other methods. The third gap will be addressed by performing an EDA analysis by plotting histogram and violin plot for the dataset. The EDA analysis will show the trend in customers' traits about their spending behavior.

The remainder of the paper is drafted in the following format. Section 2 of the paper discusses the case study and the assumptions considered for solving the problem. Section 3 briefly describes the preliminary concept of the methodology used for analyzing the dataset. Section 4 of the paper describes the result obtained after applying the proposed methodology in the case study and validates the proposed model. Finally the paper is concluded in section 5

2. Case study

In this section of the paper, a brief description of the case study along with the different assumptions, dataset and data preprocessing is discussed.

2.1. The problem

The problem of CS in malls is about segregating the customers into groups based on common characteristics with the aim to market and sell products to each group of customers effectively and appropriately. Common characteristics in CS can guide how a company markets to individual segments and what products or services it promotes to them.

In a mall, the main aim is to arrange the products in sections in such a way that each segment is primarily focussed to a certain group of customers. By segmenting the products according to the customers' needs, the malls aim at increasing their profit as well as their reputation among their customers. The CS also helps the mall to answer the following questions:

a) Will a product launched be accepted by the customers?
b) Which group of customers will mostly accept the product?
c) What will be the demand and how much to be ordered from the manufacturers?

Segmentation allows marketers to better tailor their marketing efforts to various audience subsets. Those efforts can relate to both communications and product development. Specifically, segmentation helps a mall:

I. Create and communicate targeted marketing messages that will resonate with specific groups of customers, but not with others (who will receive messages tailored to their needs and interests, instead).

II. Select the best communication channel for the segment, which might be email, social media posts, radio advertising, or another approach, depending on the segment.
III. Identify ways to improve products or new product or service opportunities.
IV. Establish better customer relationships.
V. Test pricing options.
VI. Focus on the most profitable customers.
VII. Improve customer service.
VIII. Upsell and cross-sell other products and services.

CS requires a mall to gather specific information (data) about customers and analyze it to identify patterns that can be used to create segments. Some of that can be gathered from purchasing information such as job title, age, sex, geography, products purchased. Some of it might be gleaned from how the customer entered your system. For example, an online marketer working from an opt-in email list might segment marketing messages according to the opt-in offer that attracted the customer. Other information, however, including consumer demographics such as age and marital status, will need to be acquired in other ways.

Typical information-gathering methods include:

I. Face-to-face or telephone interviews
II. Surveys
III. General research using published information about market categories
IV. Focus groups

CS can be practiced by all businesses regardless of size or industry and whether they sell online or in person. It begins with gathering and analyzing data and ends with acting on the information gathered in a way that is appropriate and effective.

2.2. The dataset

In the present study, the dataset used for CS using ML models comprises information such as customer id, date of birth, gender, annual income in k$ and amount of money spent on product purchasing. The customer id is a unique identification number given to each customer. The mall data acquisition system keeps a track of the date of birth of each customer in order to know their age and also sends gift cards or surprise offers to the customers on their birth month. By doing so the mall strengthens the relationship with the customers as well as increases their reputation among them. The dataset also comprises the annual income of the customer and the amount of money spent in the mall. This dataset is used to give an idea on the money spent and the behavior of the customer.

3. Methodology

This section of the paper briefly describes the preliminary concept required to analyze the data and subsequent development of the ML models. The ML models developed for the process of CS fall under the category of unsupervised learning (UL). Initially, recalling the definition and concept of UL.

3.1. Unsupervised learning (UL)

In the research domain of artificial intelligence (AI), UL refers to the learning algorithms that identify the patterns in datasets that are neither classified nor labeled [26]. In the field of data science, the training data comprises a set of input vectors without any corresponding target values. The goal of the UL algorithms is to segregate the similar data and group them into clusters based on the pattern identified by the algorithms. These UL algorithms are termed as clustering.

3.2. Clustering

Clustering is a UL technique with the aim of finding a structure in a collection of unlabeled data. A loose definition of clustering is the process of organizing objects into groups whose members are similar in some way. A cluster is therefore a collection of objects which are "similar" between them and are "dissimilar" to the objects belonging to other clusters. Cluster in pictorial form is shown in figure (1).

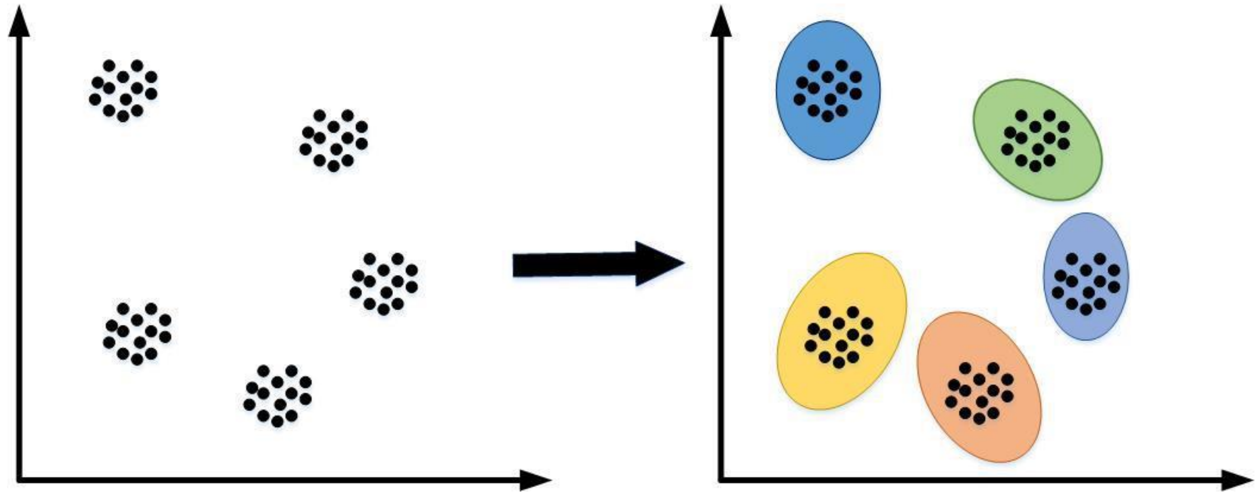


Figure (1): Visualization of cluster

Clustering can be broadly divided into four categories, namely Centroid based clustering, density based clustering, distribution based clustering and hierarchical based clustering [27]. The Centroid-based clustering organizes the data into non-hierarchical clusters. Centroid-based algorithms are efficient but sensitive to initial conditions and outliers [28]. Density-based clustering connects areas of high example density into clusters. This allows for arbitrary-shaped distributions as long as dense areas can be connected. These algorithms have difficulty with data of varying densities and high dimensions. Further, by design, these algorithms do not assign outliers to clusters [29]. Distribution based clustering approach assumes data is composed of distributions. The algorithm is based on the assumption that as distance from the distribution's center increases, the probability that a point belongs to the distribution decreases [30]. Hierarchical clustering creates a tree of clusters. Hierarchical clustering is well suited to hierarchical data, such as taxonomies [31]. Out of all the clustering algorithms, the problem of CS is mostly analyzed using centroid and density-based clustering techniques.

3.2.1 Centroid-based clustering technique

k-means is the most widely-used centroid-based clustering algorithm. The k-means algorithm identifies the k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. The 'means' in the k-means refers to averaging of the data that is used to find the centroid [32]. The steps of k-means algorithms are as follows [32]:

a. Randomly choose the number of centroids (k).
b. Assigned all the data points to the closest centroid.
c. Recomput~~ed~~ the centroids of newly formed clusters.
d. Repeated the steps (b) and (c) until one of the following three conditions are satisfied:
    i. No change in the centroids of newly formed clusters
    ii. Points remain in the same cluster
    iii. Maximum number of iterations are reached

3.2.2 Density-based clustering technique

As already defined in section 3.2 about density-based clustering, the most widely used algorithm is density-based spatial clustering of applications with noise (DBSCAN) method. The DBSCAN algorithm is based on this intuitive notion of “clusters” and “noise”. The basic concept of the algorithm is that for each point of a cluster, the neighborhood of a given radius has to contain at least a minimum number of points [33]. The steps for DBSCAN are as follows:

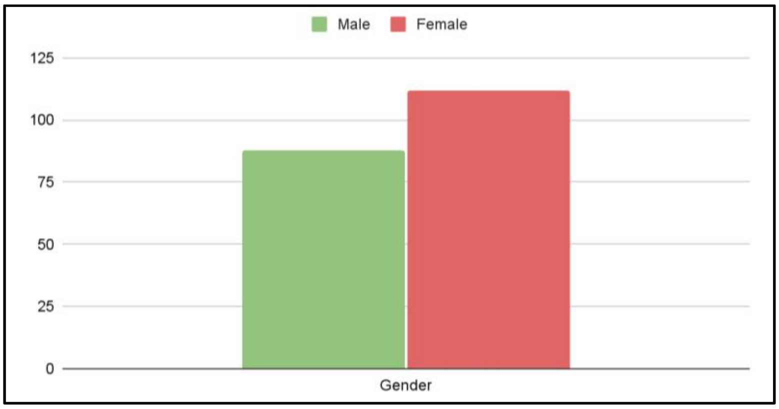
a. Classifying the points as core, boundary and noise points
b. Elimination of the noise points.
c. Assigning a core point to a cluster.
d. Color all the density connected points of a core point.
e. Color boundary points according to the nearest core point.

4. Results and discussion

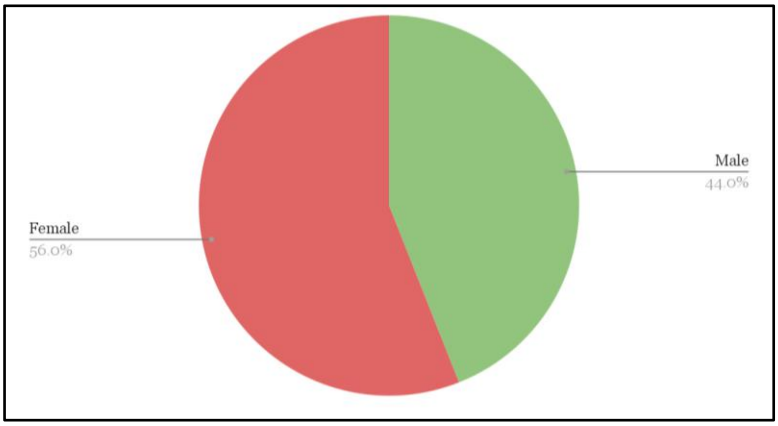
In this section of the paper, the results obtained from analyzing the datasets and a brief discussion is presented.

4.1. Exploratory data analysis (EDA)

An exploratory analysis of the data collected for the study is conducted that shows the trend of the customers’ traits about spending behavior. Figure (2) shows the gender group breakdown as a pie-chart of the collected dataset. In figure (3.a) and figure (3.b), the age of the customers for the collected data is shown in the form of a bar-chart as well as kdeplot and violin plot respectively. In figure (4.a) and figure (4.b) the annual income distribution of the customers for the collected data is shown in the form of a bar-chart as well as kdeplot and violin plot respectively. In figure (5.a) and figure (5.b), the spending score distribution of the customers for the collected data is shown in the form of a bar-chart as well as kdeplot and violin plot respectively.



(a)



(b)

Figure (2): (a) Bar-chart and (b) Pie-chart showing the gender group breakdown of the collected data

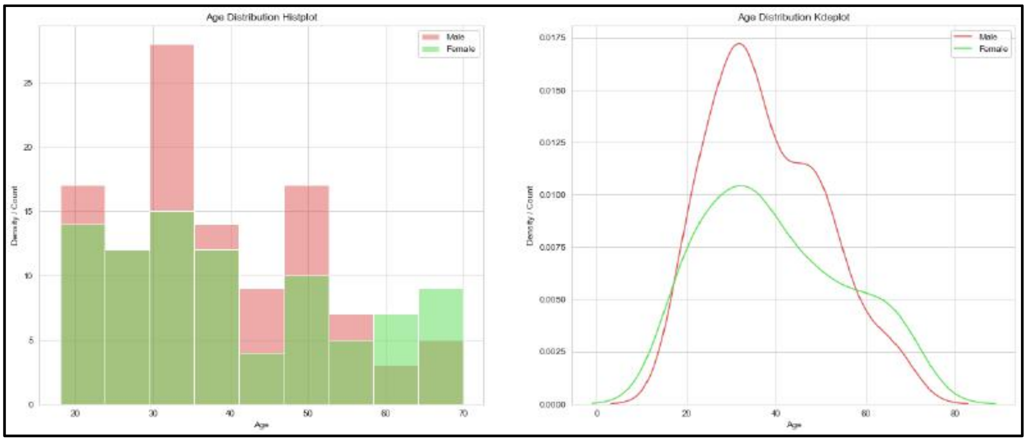


Figure (3.a): Bar-chart and kdeplot showing the age distribution of the collected data

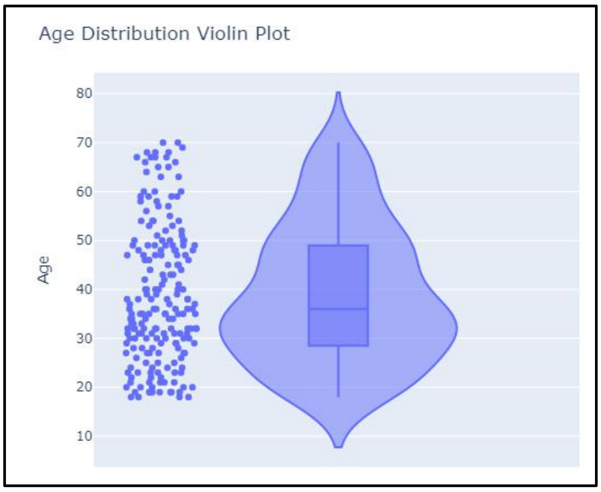


Figure (3.b): Violin plot showing the age distribution of the collected data

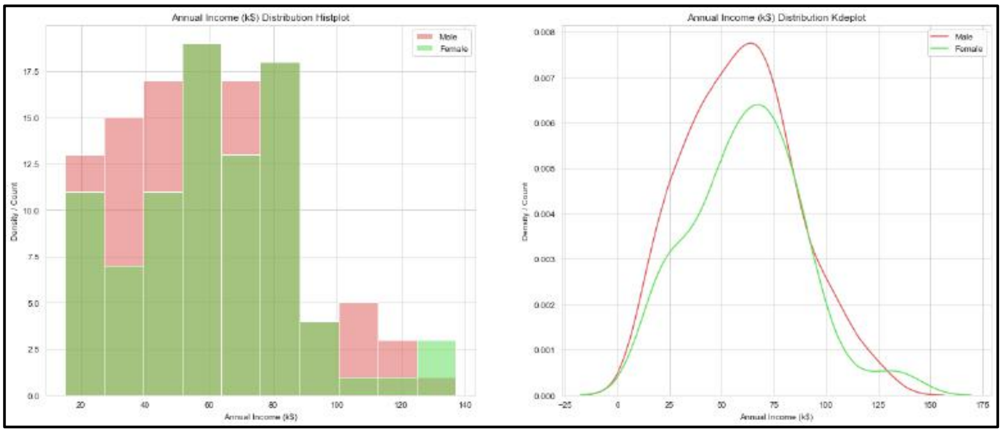


Figure (4.a): Bar-chart and kdeplot showing the annual income distribution of the collected data

*i*Journals: International Journal of Software & Hardware Research in Engineering (IJSHRE)
ISSN-2347-4890
Volume 10 Issue 6 June 2022

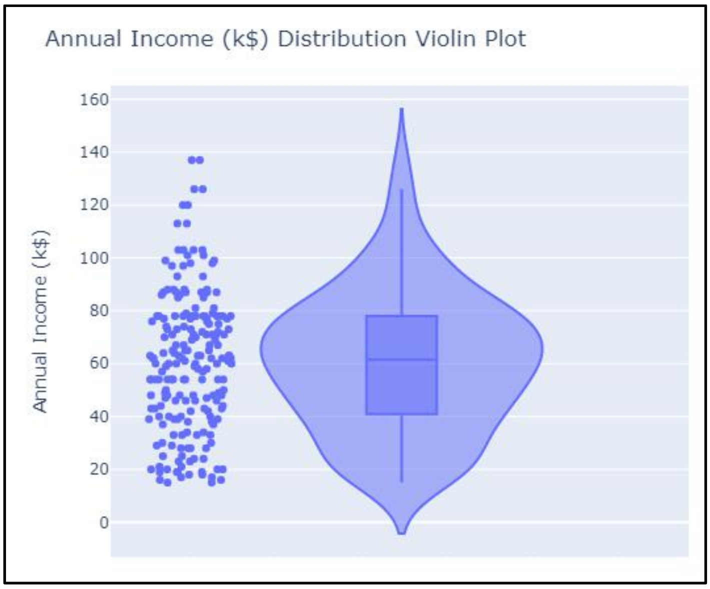


Figure (4.b): Violinplot showing the annual income distribution of the collected data

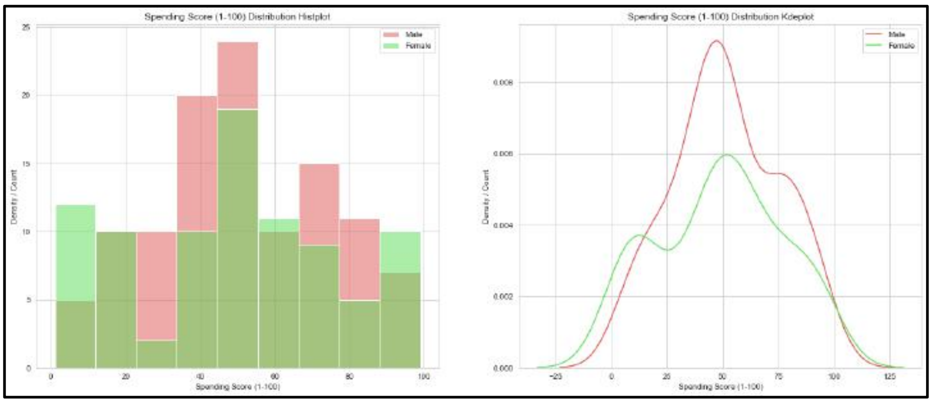


Figure (5.a): Bar-chart and kdeplot showing the spending distribution of the collected data

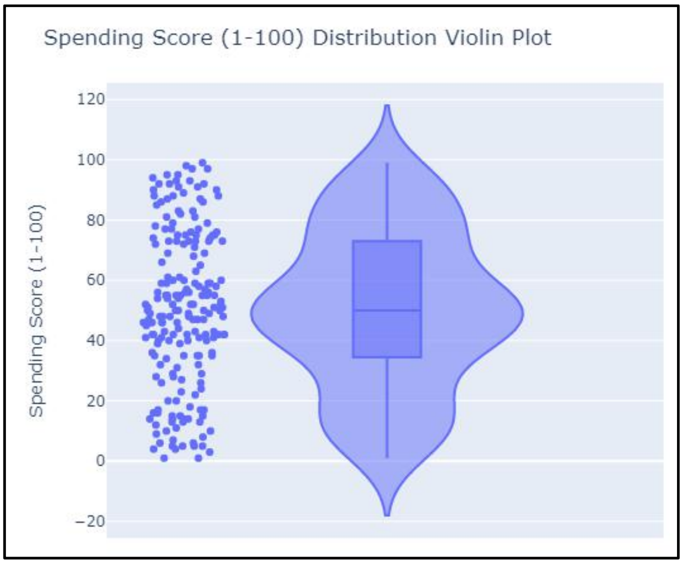


Figure (5.b): Violin plot showing the spending distribution of the collected data

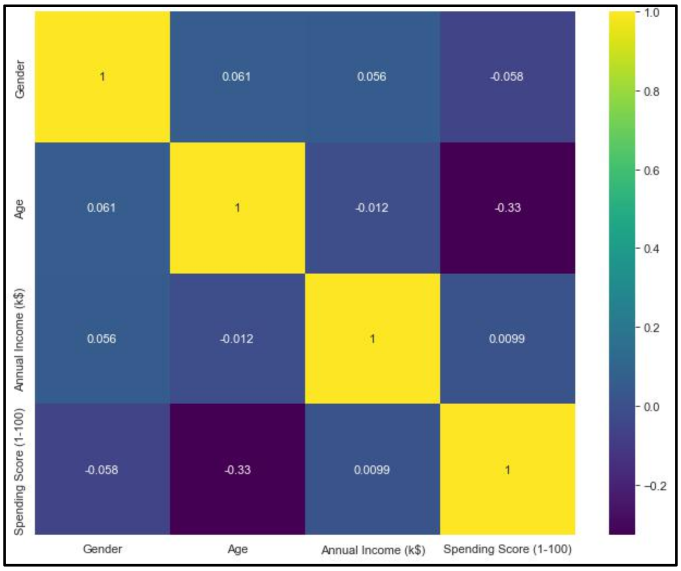


Figure (6): Correlation between different factors

Some of the important observations made from EDA are as follows:

a. From figure 2(a) and 2(b), it can be concluded that more female customers visit the mall than the male customers.
b. From figure (3.a), it is evident that more aged female customers visit the mall than the male customers.
c. However, the age group of the customers visiting the mall ranges from 18 to 70 years, yet the majority of the customers are below 40 years. This can be observed from figure (3.b)
d. From figure (4.a), it is observed that the female customers of the mall who earn the least and also the highest among all the customers.
e. From figure (4.b), it can be concluded that it is visited by customers earning $20,000 to $140,000 per year. So there should be products and services that can attract customers in the income group. It can also be observed that the majority of the customers earn less than $80,000 per year.
f. From figure (5.a), it is observed that the spending score of the customers varies from 0 to 100. The female customers spend the least as well as the most in comparison to the male customers.
g. From figure (5.b), the majority of the customers' spending score lies in the range of 40 - 60. Above that the spending attitude is somewhat uniform for all the customers irrespective of their gender, age and annual income.
h. Figure (6), shows the correlation between different factors. Some of the points observed from figure (6) are:
    i. The lower age group customers earn and spend more than the higher age group.
    ii. The female customers spend more than the male customers.
    iii. The customers who earn more also spend more.

4.2. Results from clustering algorithms

From the EDA conducted in section 4.1, it can be observed that there are several groupings. The aim of this section is to compute the number of ideal groups so that a common strategy could be made for each cluster.

4.2.1. Clustering by k-means algorithm

In k-means clustering, the ideal number of clusters can be computed using the Elbow method. The steps for the Elbow method are as follows:

a. Performed k-means clustering for 1 to 20 clusters.
b. Computed the sum of square distances from each point of each cluster to the centroid.

c. Computed the mean square (MS) value for each instance.
d. Plotted the MS-value for each instance.
e. The point after which there is no sudden change in the MS-value within clusters is chosen as the ideal number of clusters.

Following the above steps, it is observed that the ideal number of the clusters for the collected dataset is 5. The plot for MS-value for each instance from 1 to 10 clusters is shown in figure (7).

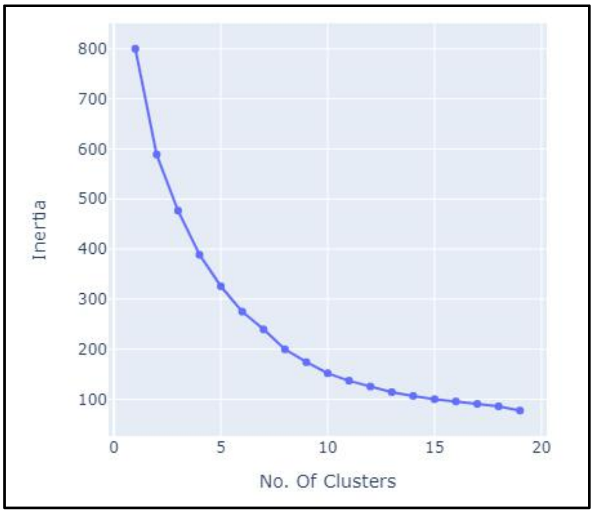


Figure (7): Plot showing the MS-value within clusters from 1 to 10

By k-means clustering algorithm, the dataset is divided into 5 clusters namely customers with Moderate Income - Moderate Spending Score (represented by purple color), High Income - High Spending Score (represented by pink color), High Income - Low Spending Score (represented by black color), Low Income - High Spending Score (represented by orange color) and Low Income - Low Spending Score (represented by white color). Figure (8) shows the clustering by k-means algorithm.

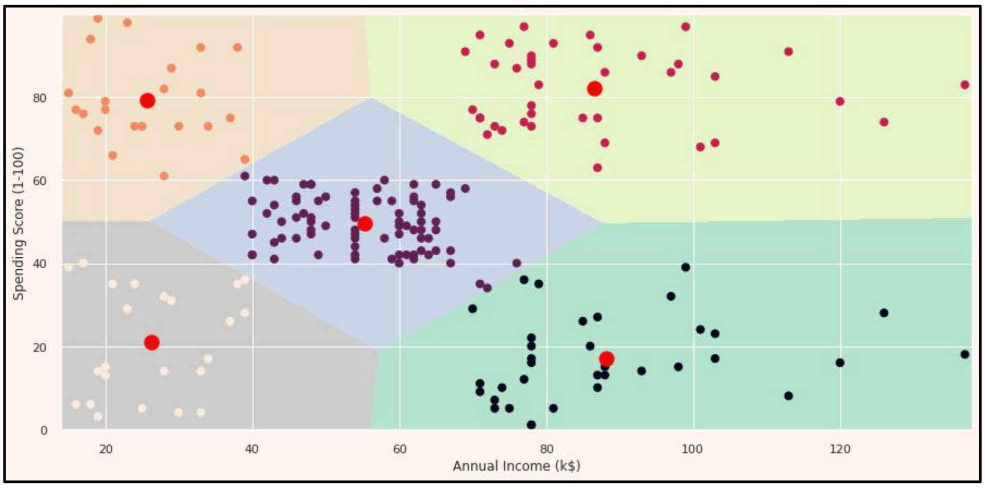


Figure (8): Clustering of the collected data by k-means

4.3. Results from DBSCAN

By DBSCAN clustering algorithm, the dataset is divided into 3 clusters namely customers with Low and Moderate annual income (represented by blue color), High Income - High Spending Score (represented by red color) and High Income - Low Spending Score (represented by green color). The algorithm has also identified certain noise data in the datasets which are represented by black color in the figure (9). The noise from the context of the problem of CS refers to those customers ~~for~~ who cannot be placed in any clusters.

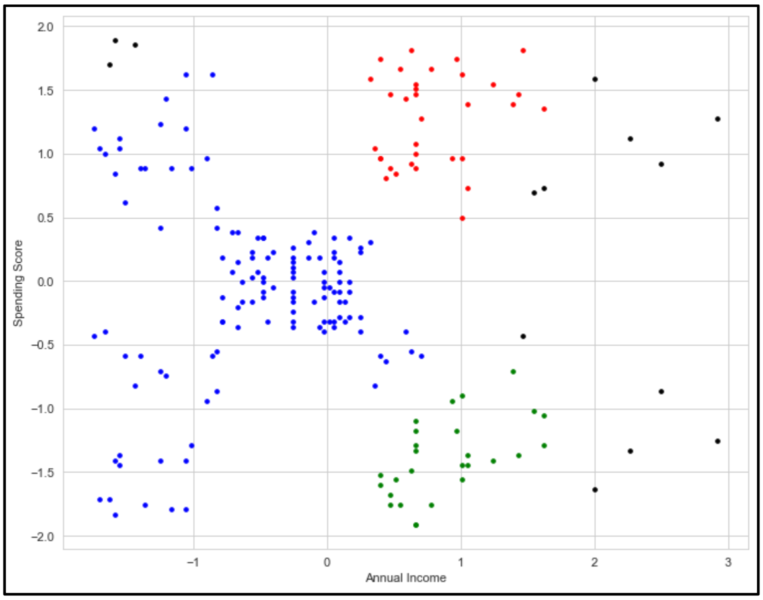


Figure (9): Clustering of the collected data by DBSCAN

5. Conclusion

The comprehensive intention of the present work is to do CS in malls using ML algorithms. The study primarily focuses on the grouping of the customers on the basis of their annual income and spending score. To meet the needs of the customers visiting the mall, the managers must be able to understand the needs of the customers by studying their shopping habits and maintain regular interactions with customers that can make them feel comfortable. To achieve these objectives, the ML algorithm is implemented to do the CS.

The problem of CS based on annual income and customers' spending behavior is a case of UL. The data are clustered using k-means and DBSCAN clustering techniques. The optimal number of clusters identified for the collected data is obtained from the Elbow algorithm. The Elbow method computed 5 clusters which are named as Moderate Income - Moderate Spending Score, High Income - High Spending Score, High Income - Low Spending Score, Low Income - High Spending Score and Low Income - Low Spending Score. However, the DBSCAN algorithm grouped the data into three clusters along with some noise data. The noise from the context of the problem of CS refers to those customers ~~for~~ who cannot be placed in any clusters. Since the noise data are also engulfed by the k-means algorithm, hence the result obtained from the k-means is the optimal.

This research proves that it is possible to implement ML in the CS problem. However, assuming ML algorithms to perform with absolute accuracy is still an exaggeration of the claim because the data collected are structured and obviously depend on human behavior. Although the claim that the result obtained from the UL algorithms is overhyped still it is safer from the management point of view in determining strategy and tactics to increase the sale and profit of the mall by targeting all the section of customers.. However, this does not close the answer that its application failed as the fact that the results obtained from the study can be arguably appropriate for use. The application of UL in this study may open up the potential for other applications in the same industry.

Acknowledgement

We would like to express my heartfelt gratitude to all the tutors and mentors of On My Own Technology pvt. ltd. for extending their help in carrying out the particular project. It is because of their help that we are able to conduct the research. We shall remain ever grateful for their help and generosity.

Conflict of interest

The authors would like to declare that no fundings in any form is received for carrying out this research work.

References

1. Kansal, T., Bahuguna, S., Singh, V., & Choudhury, T. (2018, December). Customer segmentation using K-means clustering. In 2018 international conference on computational techniques, electronics and mechanical systems (CTEMS) (pp. 135-139). IEEE.
2. Pradana, M. G., & Ha, H. T. (2021). Maximising strategy improvement in mall customer segmentation using k-means clustering. Journal of Applied Data Sciences, 2(1), 19-25.
3. Calvo-Porral, C., & Lévy-Mangin, J. P. (2019). Profiling shopping mall customers during hard times. Journal of Retailing and Consumer Services, 48, 238-246.
4. Can, G. F., Kurtulmusoglu, F. B., & Atalay, K. D. (2016). A case study on shopping malls attributes for young consumers. Young Consumers.
5. Calvo-Porral, C., & Lévy-Mangín, J. P. (2018). Pull factors of the shopping malls: an empirical study. International Journal of Retail & Distribution Management.
6. Ubeja, S., & Kusuma, N. P. (2014). A Study of Sales Promotion Mix on Customer Satisfaction With Reference to Shopping Malls in Indore. Global Journal of Finance and Management, 6(3), 245-252.
7. Pentecost, R., Donoghue, S., & Thaichon, P. (2019). Emerging from my youth–intra-cohort segmentation. International Journal of Retail & Distribution Management.
8. Katrodia, A., Naude, M. J., & Soni, S. (2018). Consumer buying behavior at shopping malls: Does gender matter?. Journal of Economics and Behavioral Studies, 10(1 (J)), 125-134.
9. Koksal, H. (2019). Shopping motives, mall attractiveness, and visiting patterns in shopping malls in the Middle east: a segmentation approach. Contemporary Management Research, 15(1), 1-23.
10. Makgopa, S. S. (2018). Determining shopping malls customers' satisfaction and loyalty. Journal of Business and Retail Management Research, 13(1).
11. Reimers, V. (2013). Convenience for the car-borne shopper: are malls and shopping strips driving customers away?. Transportation Research Part A: Policy and Practice, 49, 35-47.
12. Makgopa, S. (2016). Determining consumers' reasons for visiting shopping malls. Innovative Marketing, 12(2), 22-27.
13. Ameen, N., Tarhini, A., Shah, M., & Madichie, N. O. (2020). Going with the flow: smart shopping malls and omnichannel retailing. Journal of Services Marketing.
14. Jogee, R., & Callaghan, C. W. (2014). A segmentation of mall shopping motivations in the South African Gauteng provincial context. Mediterranean Journal of Social Sciences, 5(9), 41.
15. MONDAL, S., MALL, M., MISHRA, U. S., & SAHOO, K. (2017). Investigating the factors affecting customer purchase activity in retail stores. Revista Espacios, 38(57).
16. Oztaysi, B., Gurbuz, T., Albayrak, E., & Kahraman, C. (2016). Target Marketing Strategy Determination for Shopping Malls Using Fuzzy ANP. Journal of Multiple-Valued Logic & Soft Computing, 27.
17. Singh, D. P. (2018). Integration of materialism with shopping motivations: motivations based profile of Indian mall shoppers. Journal of Asia Business Studies.
18. Prashar, S., Singh, H., Parsad, C., & Vijay, T. S. (2019). Segmenting Indian shoppers on mall attractiveness factors. International Journal of Services Technology and Management, 25(1), 18-35.
19. Kumar, A., Kashyap, A. K., & Bhagwat, S. (2021). An exploratory analysis of mall attractive dimensions from the perspective of tier-II cities customers. International Journal of Business Innovation and Research, 25(1), 78-93.
20. Pare, V., & Pourazad, N. (2017). The Big Bazaar: an examination of Indian shopping mall behaviour and demographic differences. Asia Pacific Journal of Marketing and Logistics.
21. Mehta, D., & Chugan, P. K. (2013). The impact of visual merchandising on impulse buying behavior of consumer: A case from Central Mall of Ahmedabad India. Universal Journal of Management, 1(2), 76-8.

22. Yousaf, A., Arora, N., & Gupta, A. (2015). Determinants of customer satisfaction in context of retail malls: Empirical evidence from India. Indian journal of marketing, 45(11), 42-56.
23. Prashar, S., Gupta, S., Singh, H., Vijay, T. S., & Parsad, C. (2016). CLUSTERING SHOPPERS BY MALL EXPERIENCE FOR EMERGING INDIAN CITY. Asian Academy of Management Journal, 21(2).
24. Prashar, S., Parsad, C., & Vijay, T. S. (2017). Segmenting young Indian impulsive shoppers. Journal of International Consumer Marketing, 29(1), 35-47.
25. Khurana, T., & Dwivedi, S. (2017). Customer Satisfaction Towards Mall Attributes in Shopping Malls of Udaipur. International Journal of Environment, Ecology, Family and Urban Studies (IJEEFUS), 7, 25-28.
26. Berry, M. W., Mohamed, A., & Yap, B. W. (Eds.). (2019). Supervised and unsupervised learning for data science. Springer Nature.
27. Xu, D., & Tian, Y. (2015). A comprehensive survey of clustering algorithms. Annals of Data Science, 2(2), 165-193.
28. Uppada, S. K. (2014). Centroid based clustering algorithms—A clarion study. International Journal of Computer Science and Information Technologies, 5(6), 7309-7313.
29. Kriegel, H. P., Kröger, P., Sander, J., & Zimek, A. (2011). Density-based clustering. Wiley interdisciplinary reviews: data mining and knowledge discovery, 1(3), 231-240.
30. Yu, Z., Zhu, X., Wong, H. S., You, J., Zhang, J., & Han, G. (2016). Distribution-based cluster structure selection. IEEE transactions on cybernetics, 47(11), 3554-3567.
31. Hassan, S. I., Samad, A., Ahmad, O., & Alam, A. (2020). Partitioning and hierarchical based clustering: a comparative empirical assessment on internal and external indices, accuracy, and time. International Journal of Information Technology, 12(4), 1377-1384.
32. Likas, A., Vlassis, N., & Verbeek, J. J. (2003). The global k-means clustering algorithm. Pattern recognition, 36(2), 451-461.
33. Khan, K., Rehman, S. U., Aziz, K., Fong, S., & Sarasvady, S. (2014, February). DBSCAN: Past, present and future. In The fifth international conference on the applications of digital information and web technologies (ICADIWT 2014) (pp. 232-238). IEEE.