

## Customer Segmentation

Pradnya R. Kenjale\*, Snehal M. Bhosale\*, Gaurav B. Kadam\*,

Aniket S. Kumbhar\*, Prof. Mhaske V. D.\*\*,

\*(UG Students, Computer Engineering, SVPM'S College of Engineering Malegaon, Baramati  
Email: [pradnyakenjale4@gmail.com](mailto:pradnyakenjale4@gmail.com), [snehalbhosale0309@gmail.com](mailto:snehalbhosale0309@gmail.com), [kadamgaurav76669@gmail.com](mailto:kadamgaurav76669@gmail.com),  
[aniketkumbhar.ak36@gmail.com](mailto:aniketkumbhar.ak36@gmail.com))

\*\* (Assistant Professor, Computer Engineering, SVPM'S College of Engineering Malegaon, Baramati  
Email: [ydmhaske@engg.svpm.org.in](mailto:ydmhaske@engg.svpm.org.in))

\*\*\*\*\*

### Abstract:

In today's world customer segmentation is very important in businesses to improve the market strategies. In e-commerce it is very important to understand the customers and identify potential customers for the business. Customer segmentation helps to analyze the customers. So, segmentation helps to group the customers based on their similar characteristics and hence it saves the time and money. With effective customer segmentation it will give the benefit for the business and profit as well. Previously various techniques were used to perform customer segmentation such as RFM, SVM. This paper focuses to do the customer segmentation in an effective way, so that we can get maximum accuracy. We are using a dataset of customers having attributes like customer ID, gender, age, annual income and spending score. So by using this dataset our system is going to use K-means and DBSCAN algorithms for doing the customer segmentation. K-means and DBSCAN are very effective clustering algorithms, so that with the help of these algorithms the segmentation will be done. It will help the business and combine for improving their market strategies.

*Keywords*—Customer segmentation, K-means, DBSCAN

\*\*\*\*\*

### I. INTRODUCTION

In today's world we are more dependent on the internet for various purposes like learning, shopping etc. E-commerce is a business which refers to the process of buying and selling of a product online using internet. Now the customers prefer online shopping than going to the market. Online shopping helps them to save a lot of time. The customer can get to see varieties of products online, so they are able to purchase the things of their wish. Customers are the most important part of the business in order to achieve the success. So for analyzing the customers and grouping them based on their similarities is called as customer segmentation. Customer segmentation is very useful for businesses

for improving their marketing strategy and for better understanding of the customers. So that they can target their customers. Every person is different from other so that customer segmentation is become an important process. Customer segmentation means it groups the customers according to the common characteristics between them. By analyzing the data of the customers businesses are able to identify the least and most profitable customers and so they can give the products of their demand and offer better service to customers. Customer segmentation helps to make a loyal relationship between the customers and the companies. It also helps businesses to enhance competitiveness in the competitive world and help to improve their marketing strategies [1].

Customer segmentation is useful technique for improving the market strategies. By effectively segmenting their customers, companies can improve their overall marketing efforts and drive greater customer loyalty and satisfaction. By segmenting customers, companies can create more targeted and effective marketing campaigns that appeal to specific groups. This report will discuss the different types of customer segmentation that companies use to better understand their customers and tailor their marketing efforts.

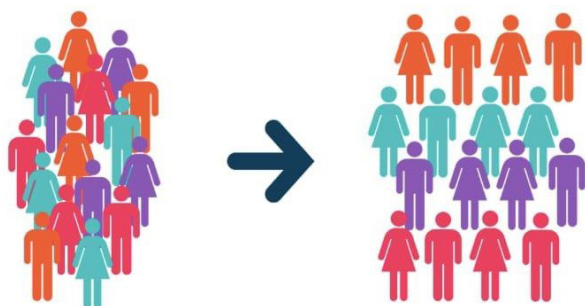


Fig. Customer segmentation

Following are the various customer segmentation types:

#### A. Demographic Segmentation

Market segmentation is a crucial strategy in marketing, and demographic segmentation is one of the most commonly used techniques. Age, occupation, gender, and marital status are various demographic factors. For instance, a cosmetic company may use demographic segmentation to target women in the age group of 18-35 years who are more likely to be interested in beauty products, thus increasing the effectiveness of their marketing campaigns and ultimately boosting sales.

#### B. Geographic Segmentation

Geographic segmentation divides customers based on their location. This type of segmentation is useful for companies that have a physical presence in a specific region or want to tailor their marketing campaigns to a particular location. For example, a coffee shop chain might target customers in urban areas where there is a high demand for coffee.

#### C. Behavioural Segmentation

Behavioural segmentation divides customers based on their behaviour, such as purchasing habits, brand loyalty, and product usage. This type of segmentation is useful for companies that want to understand their customers' preferences and purchasing patterns. For example, a company that sells pet food might target customers who buy premium pet food products, as these customers are more likely to be loyal and willing to pay more for high-quality products.

#### D. Psychographic Segmentation

Psychographic segmentation divides customers according to their personality traits, values, and beliefs. This segmentation helps the companies that want to understand their customers' motivations and desires.

To gain profit and improve market strategies of business, we are going to develop a customer segmentation system which will help the business to identify target customers. They are also able to identify most and least profitable customers for the company. So, it will be very helpful.

## II. LITERATURE SURVEY

In today's world, continuous growth and development of e-commerce, the customer, product, and service to the customer become very important. So, customer segmentation to find the target customers becomes very important. The research of this customer segmentation were constructed on three-dimensional customer segmentation, such as customer lifetime value, customer activity, and customer satisfaction [3]. So, that the customer can be divided into the different groups correctly. So with the help of it e-commerce website will do the customer segmentation. There are various machine algorithms are used for customer segmentation.

The clustering algorithms such as K-means, agglomerative, and mean shift have been implemented to segment the customers into different groups [7]. Clustering is very effective for the implementation of customer segmentation. Clustering means finding clusters on the unlabelled dataset. Clustering is the technique which comes under the unsupervised learning.

In the current era, with the widespread use of e-commerce websites for shopping, customers share their valuable information with the website's servers. This information is crucial in evaluating the business' performance, and predictive analytic techniques are applied to analyze the data and anticipate future customer behaviour. The analysis considers a variety of factors, such as the products purchased by the customer, the quality of the products, and the price of the items, which are then utilized to manage inventory levels and enhance customer satisfaction.

To facilitate this analysis, Support Vector Machines (SVMs) are utilized, which leverage multidimensional hyperplanes to classify the data. The SVM technique was developed based on neural network principles and has proven to be a reliable and accurate method for classification tasks. Additionally, the SVM model utilizes a sigmoid kernel function to help identify patterns and trends within the data. By employing SVM techniques, businesses can better understand customer behavior and can reach to the changing demands of their clientele [6].

In this world full of competition, every online business wants to know their customer's true value and loyalty using their data. Customer Segmentation helps to increase customer satisfaction and as well as revenue of the business. For this task every company is establishing Customer Relation Management (CRM). The loyalty of a company's customers is closely tied to their past purchasing behaviour. In order to better understand and measure customer loyalty, various customer lifetime value models have been developed and used in research. One such model is RFM, which stands for Recency, Frequency, and Monetary value. This model is used to analyze customer behaviour based on the recency of their purchases, frequency of their purchases, and monetary value of their purchases. By using the RFM model, companies can gain insights into which customers are most valuable to their business and develop strategies to increase customer loyalty and retention. To calculate the value of RFM, Fuzzy-AHP method is used [4]. For predicting the

retention of customer and profit the random forest and regression forest techniques are used [9]. For making successful customer relationship strategy it is important to know the class of customer which can be calculated using Customer Lifetime Value (CLV). Random forest algorithm is used for this purpose [10].

Table 1: Efficiency of Different Algorithms

Sr. No.	Title	Algorithm	Efficiency
1.	Review on customer segmentation technique on e-commerce	Hierarchical clustering, Affinity propagation clustering	Low
2.	Predicting customer behaviour in online shopping using SVM classifier	Support vector machine (SVM)	Medium
3.	Predicting customer retention and profitability by using random forests and regression forest techniques.	Random forest and regression forest	Good
4.	Predicting Customer Class using Customer Lifetime Value with Random Forest Algorithm.	Random forest, adaboost	Good

### III. PROPOSED SYSTEM

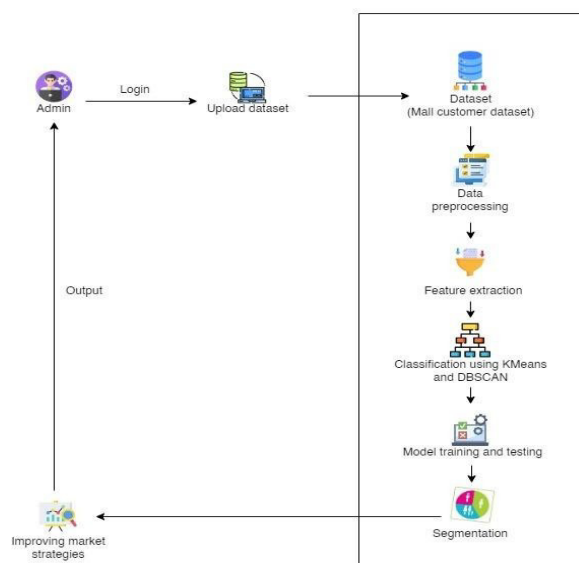


Fig. Proposed system

The proposed works in the following manner:

#### A. Dataset

For implementing customer segmentation system we are going to use a dataset of customers. The dataset will contain the different attributes of customers such as customer ID, gender, age, annual income and spending score. So by using this information of customers we are going to provide this data to our system for further processing. The admin will upload the dataset of customers to the system.

#### B. Data preprocessing

Data preprocessing is a very important step, it is important to improve data quality. It has steps such as:

1. Data cleaning: It includes filling missing values, removing noisy data, resolving inconsistency and removing outliers. We can fill missing values using manually or by calculating mean etc. Noisy data can be removed using binning, regression and clustering.

2. Data Integration: Data integration means merging data from various heterogeneous sources to attain meaningful data.
3. Data transformation: It is the process of converting data from one format to another. It can be done using generalization, normalization, attribute selection and aggregation methods.
4. Data reduction: Data reduction is used to reduce the size of a dataset but also preserve the important information. Efficiently managing large volumes of data is crucial in today's information-driven world, and there are various techniques that can be employed to achieve this goal.

#### C. Feature extraction

In feature extraction it reduces redundant data and will only extract the useful features. Feature extraction techniques areas follows:

1. PCA (Principal Component Analysis): PCA is a popular linear technique that works by identifying the most important combinations of input features that best summarize the original data distribution. By reducing the number of dimensions, PCA helps to simplify data sets and facilitate more efficient analysis [11].
2. ICA (Independent Component Analysis): ICA, on the other hand, is another linear technique that focuses on identifying independent components within the input data. By removing unnecessary noise and identifying independent features, ICA enables more accurate analysis and modeling of complex datasets [11].
3. LDA (Linear Discriminant Analysis): LDA, meanwhile, is a supervised learning method. By identifying the most important features that contribute to class separation, LDA enables more accurate classification and prediction of datasets [11].

#### D. Classification using Kmeans and DBSCAN

### **Kmeans:**

K-means is a widely used unsupervised machine learning algorithm that enables data scientists to group unlabelled datasets into distinct clusters. This powerful clustering technique has several key steps that enable it to identify meaningful patterns and relationships within large and complex data sets. The first step in the K-means is that it finds the number of clusters from the given data. This can be achieved through a variety of methods, such as the elbow method or silhouette analysis, which enable data scientists to identify the optimal number of clusters for a given data set. When the number of clusters are decided, K-means algorithm selects an initial set of centroids at random and assigns each data point to the centroid which is near. This process continues until all data points have been assigned to a cluster. Next, the algorithm calculates the mean of all objects assigned to each centroid and reassigns data points to the nearest centroid based on their distance from the updated centroid. This process continues until no change is made when we assign data point to cluster.

Overall, the K-means algorithm is a powerful tool for clustering unlabelled datasets and identifying meaningful patterns and relationships within complex data.

### **DBSCAN:**

DBSCAN is a powerful density-based clustering algorithm that is designed to identify clusters of varying shapes and sizes within large datasets, even in the presence of noise and outliers. DBSCAN automatically detects the optimal number of clusters based on the density of the data. DBSCAN is capable of differentiating between high-density and low-density clusters, by identifying points that are closely packed together and have a higher density

than points outside of the cluster. Points that do not belong to any cluster are considered to be noise points, and are often removed from the dataset during pre-processing. Benefit of DBSCAN is it is able to manage data sets with varying densities and irregular shapes. It is particularly useful in applications where the data is inherently noisy or contains outliers, as it can effectively separate them from the main clusters. DBSCAN is widely used in a variety of domains such as image processing, social network analysis, and anomaly detection, among others.

### **E. Output**

The output of this will be these segmented result of the customers. We are able to identify the target customers using the result of segmentation. So, the overall marketing strategy of the business will get improved.

## **IV. RESULT ANALYSIS**

In this paper we have discussed different papers. There are various algorithms are discussed in different papers for customer segmentation such as hierarchical clustering, SVM (Support Vector Machine), random forest, AdaBoost etc. These all algorithms are used for performing customer segmentation. By analysing these algorithms we can determine that the SVM and random forest algorithms are good for the customer segmentation. They are efficient also. But K-means and DBSCAN algorithms are more efficient than these algorithms. So we have used K-means and DBSCAN algorithms.

## **V. ADVANTAGES**

1. Help to increase the profit of the business
2. To identify least and most profitable customers.
3. For improving marketing strategies.



4. For identifying the target customers.

## VI. CONCLUSION

In this review paper various techniques are discussed for the customer segmentation. In this way we can conclude that effective customer segmentation is a crucial aspect of any successful business strategy, as it enables companies to better understand and cater to the needs of their diverse customer base. One effective way to segment customers is using K-means clustering, a powerful machine learning technique that groups customers based on similarities in their purchasing behavior, demographics, and other relevant factors. SVM is also used for segmenting the customers. These algorithms are efficient but we can also extend for getting more efficient customer segmentation.

## VII. ACKNOWLEDGMENT

We extend our heartfelt gratitude to our project guide, Prof. Mhaske V. D., for her invaluable guidance and support throughout the duration of our project. We also express our sincere appreciation to the faculty members of the computer engineering department at SVPM's College of Engineering, Malegaon Bk, Bara mati, Pune 413115, for granting us the opportunity to conduct our project work and for their continuous encouragement and assistance in all aspects of our research.

## VIII. REFERENCES

[1] Sari, Juni Nurma, et al. "Review on customer segmentation technique on e-commerce." *Advanced Science Letters* 22.10 (2016): 3018-3022.  
[2] Nandapala, E. Y., & Jayasena, K. P. (2020). The practical approach in Customer segmentation by using the K-Means Algorithm. 2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS).

[3] He, X., & Li, C. (2016). The Research and Application of Customer Segmentation on E-Commerce Websites. 2016 6th International Conference on Digital Home (ICDH).  
[4] Aggarwal, A. G., & Yadav, S. (2020). Customer Segmentation Using Fuzzy-AHP and RFM Model. 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO).  
[5] Lai, X. (2009). Segmentation Study on Enterprise Customers Based on Data Mining Technology. 2009 First International Workshop on Database Technology and Applications.  
[6] Maheswari, K., & Priya, P. P. A. (2017). Predicting customer behavior in online shopping using SVM classifier. 2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS).  
[7] Kansal, T., Bahuguna, S., Singh, V., & Choudhury, T. (2018). Customer Segmentation using K-means Clustering. 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS).  
[8] Sista, R., Singh, R., Kumawat, S. K., & Dhanare, R. (2021). Techniques used by E-commerce industries for Customer analysis. 2021 International Conference on Computer Communication and Informatics (ICCCI).  
[9] Larivière, Bart, and Dirk Vanden Poel. "Predicting customer retention and profitability by using random forests and regression forest techniques." *Expert systems with applications* 29.2 (2005): 472-484.  
[10] Win, T. T., & Bo, K. S. (2020). Predicting Customer Class using Customer Lifetime Value with Random Forest Algorithm. 2020 International Conference on Advanced Information Technologies (ICAIT).  
[11] <https://towardsdatascience.com/feature-extraction-techniques-d619b56e31be>