REPORT OF *CUSTOMER SEGMENTATION USING K-MEANS CLUSTERING*

By CHETHAN

4nm19mee025@nmamit.in

ABSTRACT:

The zeitgeist of modern era is innovation, where everyone is embroiled into competition to be better than others. Today's business run on the basis of such innovation having ability to enthral the customers with the products, but with such a large raft of products leave the customers confounded, what to buy and what to not and also the companies are nonplussed about what section of customers to target to sell their products. This is where machine learning comes into play, various algorithms are applied for unravelling the hidden patterns in the data for better decision making for the future. This elude concept of which segment to target is made unequivocal by applying segmentation. The process of segmenting the customers with similar behaviours into the same segment and with different patterns into different segments is called customer segmentation. In this paper, 3 different clustering algorithms (k-Means, Agglomerative, and Meanshift) are been implemented to segment the customers and finally compare the results of clusters obtained from the algorithms. A python program has been developed and the program is been trained by applying standard scaler onto a dataset having two features of 200 training sample taken from local retail shop. Both the features are the mean of the amount of shopping by customers and average of the customer's visit into the shop annually. By applying clustering, 5 segments of cluster have been formed labelled as Careless, Careful, Standard,Target and Sensible customers. However, two new clusters emerged on applying mean shift clustering labelled as High buyers and frequent visitors and High buyers and occasional visitors.

Contents :

- ❖ Introduction 1
- ❖ Exiting Methods 1-3
- ❖ Proposed Method 3-4
- ❖ Implementation 4-5
- ❖ Results 5-6
- ❖ Conclusion 6

INTRDUCTION:

As more and more business being coming up every day, it has become significantly important for the old businesses to apply marketing strategies to stay in the market as the competition has been cut to throat. Change or die have become the simple rule of marketing in today's world. As the customer base is increasing day by day it has become challenging for the companies to cater to the needs of each and every customer, this is where Data mining serves a very important role to unravel hidden patterns stored in the company's database. Customer segmentation is one of the application of data mining which helps to segment the customers with similar patterns into similar clusters hence, making easier for the business to handle the large customer base. This segmentation can directly or indirectly influence the marketing strategy as it opens many new paths to discover like for which segment the product will be good, customising the marketing plans according to the each segment, providing discounts for a specific segment, and decipher the customer and object relationship which has been previously unknown to the company. Customer segmentation allows companies to visualise what actually the customers are buying which will prompt the companies to better serve their customers resulting in customer satisfaction, it also allows the companies to find who their target customers are and improvise their marketing tactics to generate more revenues from them.

Existing Method:

Demographic, psychographic, behavioral and geographic segmentation are considered the four main types of market segmentation, but there are also

many other strategies you can use, including numerous variations on the four main types. Here are several more methods you may want to look into.

Value segmentation: Some businesses will split up a market based on the "transactional worth" of their customers — how much they're likely to spend on their products. To determine a customer's transactional worth, you can look at previous purchase data such as how many purchases they make, how often they

make purchases and the value of the items they purchase.

Firmographic segmentation: Business-to-business (B2B) companies may use firmographic segmentation to divide up the businesses in a market. This is similar to demographic segmentation with individual consumers but instead looks at the characteristics of companies that may become customers. Examples of data to look at include industry, revenue, number of employees and location.

Generational segmentation: Businesses may segment consumers by generation and group them into categories that include Gen Z, Millenials, Generation X, Baby Boomers and the Silent Generation. These generations are believed to share certain preferences, behaviors, personality traits and beliefs. Of course, not every member of a generation is the same, but generational segmentation can give you some additional insight into your audience.

Lifestage segmentation: You can also segment your market into groups based on where they are in their lives. Going to college, getting married and having children are examples of key life events to consider.

People at different stages of life need different things. For instance, soon-to-be college students may need apartment furniture. New parents will be looking to purchase baby food.

Seasonal segmentation: Similarly to how people buy different products in different periods of their lives, people also buy different items at different times of the year. Major holidays such as Christmas and Hanukkah also significantly impact purchasing behaviors.

Proposed Method:

The Machine Learning is the most modern technique used for market segmentation. Machine learning (ML) is the study of computer algorithms that can improve automatically through experience and by the use of data. It is seen as a part of artificial intelligence.

Machine learning algorithms build a model based on sample data, known as training data, in order to make predictions or decisions without being explicitly programmed to do so.

Machine Learning techniques are broadly divided into two parts

- Supervised Machine Learning
- Unsupervised Machine Learning
-

In Supervised Machine Learning, the data is labelled and the algorithm learns from labelled training data. Examples of this method are Classification and Regression

In Unsupervised Machine Learning, we do not need to supervise the model. Such a method deals with unlabelled data. Unsupervised machine learning helps us find hidden and unknown patterns in data. Often it easier to get unlabelled data as compared to labelled data, and in such cases, we can use unsupervised machine learning to work on the data. Data, which needs categorization can be categorized with the help of unsupervised machine learning. Clustering is a type of unsupervised machine learning in which the algorithm processes our data and divided them into “clusters”. Clustering algorithms try to find natural clusters in data, the various aspects of how the algorithms to cluster data can be tuned and modified. Clustering is based on the principle that items within the same cluster must be similar to each other.

The data is grouped in such a way that related elements are close to each other. In the field of marketing, clustering can be used to identify various customer groups with existing customer data. Based on that, customers can be provided with discounts, offers, promo codes etc.

Implementation (K-Means Clustering) :

K-Means clustering is an unsupervised machine learning algorithm that divides the given data into the given number of clusters. Here, the “K” is the given number of predefined clusters, that need to be created.

It is a centroid based algorithm in which each cluster is associated with a centroid. The main idea is to reduce the distance between the data points and their respective cluster centroid.

The algorithm takes raw unlabelled data as an input and divides the dataset into clusters and the process is repeated until the best clusters are found.

K-Means is very easy and simple to implement. It is highly scalable, can be applied to both small and large datasets. There is, however, a problem with

choosing the number of clusters or K. Also, with the increase in dimensions, stability decreases. But, overall K Means is a simple and robust algorithm that makes clustering very easy.

K Means Clustering Algorithm:

- Specify number of clusters
- Initialize centroids by first shuffling the dataset and then randomly selecting K data points for the centroids without replacement.
- Keep iterating until there is no change to the centroids. i.e assignment of data points to clusters isn't changing.

Environment and tools :

- Scikit-learn
- Seaborn
- Numpy
- Pandas
- Matplotlib

Results :

Cluster 1

Customers in this cluster buy fewer items because of their minimal annual income. Generally this cluster consists of people with age above 35.

Cluster 2

People in this group spend most of the money in buying things. They are not having good annual income and buy things with parent's money or others. Most of the youngsters without their own income constitutes this group.

Cluster 3

Customers in this cluster buy a lot in apparel. They purchase the most items, and have the highest spending per item. The cluster skews slightly female. This cluster consists of the most youngsters. They can be called as Spendthrifts.

Cluster 4

This is the most populated cluster. Customers in this cluster make small numbers of purchases and spend less on average. They are typically one time buyers. The people in this cluster can be called as The Compulsive Saver or Tightwads.

Conclusion:

Considering both the centroid and summarized statistics data, you can determine the following characteristics for the model's clusters:

Principle of market segmentation is that the product and services needs of individual customers differ. Market segmentation involves the grouping customers together with the aim of better satisfying their needs whilst maintaining economies of scale. It consists of three stages and if properly executed should deliver more satisfied customers, few direct confrontations with competitors, and better designed marketing programmers. Segmentation depends the differentiation of market strategy and the achieve of goals that the main important thing with segment market don’t have achieve the target, market segmentation of different ways that the concludes are Segmentation to helps the market to pick up the market targets and that the differentiation of main thing is bases of segmentation they decided the all level of ages customer needs and wants. Segments must be identifiable so that the market can determine which consumers belong to a segment and which do not This collection refers to the degree to which a chosen segment is large enough to support profitably a separate marketing program. As was noted preciously a strategy of market segmentation is costly. Thus, one must carefully consider not only the number of customers available in a segment but also the amount of their purchasing power.

****