Customer Segmentation – ML

K. Nikhil Sai, N. Jyothi Swarupa, Uday, Satti Venkata Rama Krishna Reddy, K. Sushma Rachel

SRM University, Andhra Pradesh

(nikhilsai_kanchanapally, jyothi_nallani, kiran_dasari, Ramakrishna_satti, sushmarachel_katikitala)@srmap.edu.in

Abstract

We are living in a world where a large and vast amount of data is collected daily. Analyzing such data is an important need. In the modern era of innovation, where there is a large competition to be better than everyone, the business strategy needs to be according to the modern conditions. The world today runs on the basis of innovative ideas as there are a large number of potential customers who are confused about what to buy and what not to buy. The companies doing the business are also not able to diagnose the target potential customers.

Many algorithms are applied to identify the hidden patterns in the data for better decision-making. The concept of which customer segment to target is done using the customer segmentation process using the clustering technique. The clustering algorithm we used is the K-means algorithm which is the partitioning algorithm, to segment the customers according to similar characteristics.

I. Introduction

The large raw data that is available has resulted in the widespread use of data mining techniques in extracting meaningful and strategic information. Data mining is the process where methods are applied to extract data patterns in order to present them in a human-readable format that can be used for the purpose of decision support. They partition the data objects into groups or clusters so that objects within a cluster are similar to one another and dissimilar to objects in other clusters.

Customer Segmentation is the process of division of the customer base into several groups called customer segments such that each customer segment consists of customers who have similar characteristics. The segmentation is based on the similarity in different ways that are relevant to marketing such as gender, age, interests, and miscellaneous spending habits. Customer segmentation has importance as it includes, the ability to modify the programs of the market so that it is suitable to each customer segment, support in the business decision; identification of products associated with each customer segment, and managing the demand and supply of that product; identifying and targeting the potential customer base and predicting customer defection, providing directions in finding the solutions.

II. Literature Review

Over time, the marketable world has become more competitive, as associations similar to these have to meet the requirements and solicitations of their guests, attract new guests, and therefore ameliorate their businesses. The task of relating and meeting the requirements and conditions of every client in the business is veritably delicate. This is because guests can vary according to their requirements, wants, demographics, size, taste, and taste, features, etc.

As it is, it's a bad practice to treat all guests inversely in business. This challenge has espoused the conception of client segmentation or request segmentation, where consumers are divided into groups or parts, where members of each subcategory exhibit analogous request actions or characteristics. Consequently, Customer segmentation is the process of dividing the request into indigenous groups.

Customer segmentation gives businesses the opportunity to better adapt their advertising attempts to different audience subsets. This can help boost both product development and communication. Customer segmentation allows businesses to select the communication channels that are most suitable to reach their target audience, test different pricing decisions, create and broadcast targeted advertisements that can have maximum impact on specific sections of customers, identify areas of
improvement for existing products and service and identify the more profitable groups of customers.

Companies include billions of data about their customers, suppliers, and operations, and millions of internally connected sensors are sent to the real world on devices such as mobile phones and cars, sensing, manufacturing, and communications data.

III. Proposed Models

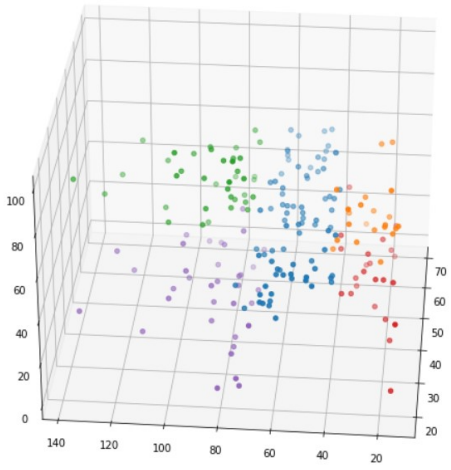
Data set Description: The dataset represents the basic data about the customers of a mall like customer ID, age, gender, annual income and spending score. The dataset consists of five columns and 200 rows. The count value of every column in the dataset has 200 which says that there are no noisy, duplicate or null values in the dataset. CustomerId, Age, Annual Income, Spending score has minimum value of 1.0000, 18.0000, 15.00000, 1.0000 respectively and max value of 200.00000 70.0000, 137.0000, 99.0000 respectively.

```
data.describe()
```

	CustomerID	Age	Annual Income (k$)	Spending Score (1-100)
count	200.000000	200.000000	200.000000	200.000000
mean	100.500000	38.850000	60.560000	50.200000
std	57.879185	13.969007	26.264721	25.823522
min	1.000000	18.000000	15.000000	1.000000
25%	50.750000	28.750000	41.500000	34.750000
50%	100.500000	36.000000	61.500000	50.000000
75%	150.250000	49.000000	78.000000	73.000000
max	200.000000	70.000000	137.000000	99.000000

Data Pre-processing: Data Pre-processing is used to transform raw data in a useful and efficient manner. Dataset can be cleaned and transformed by
using Data pre-processing techniques like data cleaning, Data transformation, data reduction etc. As the dataset used here has the count value of 200 for all the columns which represents that the dataset can be implemented directly.

IV. Results



By the above 3d plot we can see customers are divided into 5 different clusters based on their Age, Spending score and Income. The distribution of the customers into the clusters was generally good, but the standard deviations were a little high.

We think the best strategy would be to target high-income customers. The reason is that some of the high-income customers spend high money, while a significant portion of these customers spend low, there may be some things that low-spenders are not satisfied with. Improvements to be made in service and quality can increase the spending of low-income customers.

V. Conclusion

Customer segmentation is essential as it can have a positive impact on a business. It is not wise to serve all customers with the same product model, email, message or advertisement. Customers have different needs. Single approach to business may lead to less engagement and sales. Finding an optimal number of unique customer groups will help in understanding how customers differ and what they exactly want.
Customer segmentation can be the solution. It improves customer experience, increases company revenue and can also get more customers. And using machine learning we can get control over complete process. It can discover different groups and with meaningful customer base permits we can understand what exactly a customer want.
Here we create segments of customers with K-means clustering algorithm and analyze the datasets in different ways. We also visualize the dataset to know the details of the dataset and find the relation between different columns.

VI. References

Ryan Henry Papetti (2019), 'CUSTOMER SEGMENTATION ANALYSIS OF CANNABIS RETAIL DATA: A MACHINE LEARNING APPROACH'
URL : *https://core.ac.uk/download/pdf/288175101.pdf*