

CUSTOMER SEGMENTATION USING DATA SCIENCE K-MEANS CLUSTERING EXPOSYS DATA LABS

Name: Harshdeep kaur Sandhu
Institute: SGGSIET, Nanded
Internship: Data Science

ABSTRACT

This era is full of innovation, where everyone competes to be better than others. Today's business runs based on such innovation having the ability to attain the customers with the products, but 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 to decode the hidden patterns in the data for better decision making for the future.



This concept of which segment to target is well-defined by applying segmentation. The process of segmenting the customers with similar behaviors into the same segment and with different patterns into different segments is called customer segmentation.

In this project, the K-means clustering algorithm is implemented to segment the customers and finally compare the results of clusters obtained from the algorithm. By a python program, a data-sheet of 200 mall customers is turned into a well-structured segment. By k-means clustering data is sorted according to common characteristics like gender, age, annual income, and spending habits.



TABLE OF CONTENTS

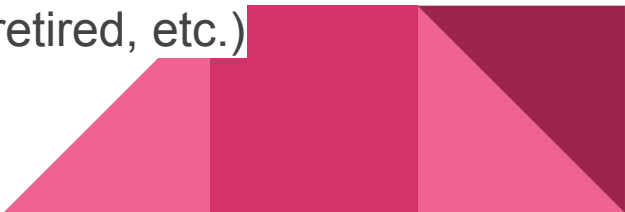
1. Introduction
2. Existing Method
3. The proposed method with Architecture
4. Methodology
5. Implementation
6. Conclusion



INTRODUCTION

Customer segmentation is the process of dividing customers into groups based on common characteristics so companies can market to each group effectively and appropriately.

In business-to-consumer marketing, companies often segment customers according to demographics that include:

- Age
 - Gender
 - Marital status
 - Location (urban, suburban, rural)
 - Life stage (single, married, divorced, empty-nester, retired, etc.)
- 

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 company:

- 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).
- Select the best communication channel for the segment, which might be email, social media posts, radio advertising, or another approach, depending on the segment.



- Identify ways to improve products or new product or service opportunities.
- Establish better customer relationships.
- Test pricing options.
- Focus on the most profitable customers.
- Improve customer service.
- Upsell and cross-sell other products and services.



EXISTING METHOD

Customer segmentation requires a company 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 – job title, geography, products purchased, for example. Some of it might be gleaned from how the customer entered your system. An online marketer working from an opt-in email list might segment marketing messages according to the opt-in offer that attracted the customer, for example.



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:

- Face-to-face or telephone interviews
- Surveys
- General research using published information about market categories
- Focus groups



THE PROPOSED METHOD WITH ARCHITECTURE

Unsupervised learning allows us to approach problems with little or no idea what our results should look like. We can derive structure from data where we don't necessarily know the effect of the variables.

We can derive this structure by clustering the data based on relationships among the variables in the data.

With unsupervised learning there is no feedback based on the prediction results.



Example:

Clustering: Take a collection of 1,000,000 different genes, and find a way to automatically group these genes into groups that are somehow similar or related by different variables, such as lifespan, location, roles, and so on.

Non-clustering: The "Cocktail Party Algorithm", allows you to find structure in a chaotic environment. (i.e. identifying individual voices and music from a mesh of sounds at a cocktail-party). 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 this project we'll use K – Means clustering algorithm to segment the cluster.

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.

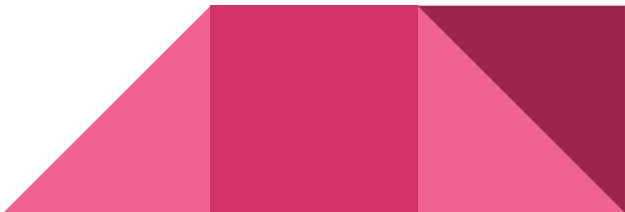


It is a centroid based algorithm in which each cluster is associated with a centroid. The main idea is to reduce the distance between the data points and their respective cluster centroid.

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.



METHODOLOGY

The data set used to implement clustering and KMeans algorithm was collected from a store of shopping mall. The data set contains 5 attributes and has 200 tuples, representing the data of 200 customers. The attributes in the data set has CustomerId, gender, age, annual income(k\$), spending score on the scale of (1-100).



IMPLEMENTATION

How does the K-Means Algorithm Work?

The working of the K-Means algorithm is explained in the below steps:

Step-1: Select the number K to decide the number of clusters.

Step-2: Select random K points or centroids.

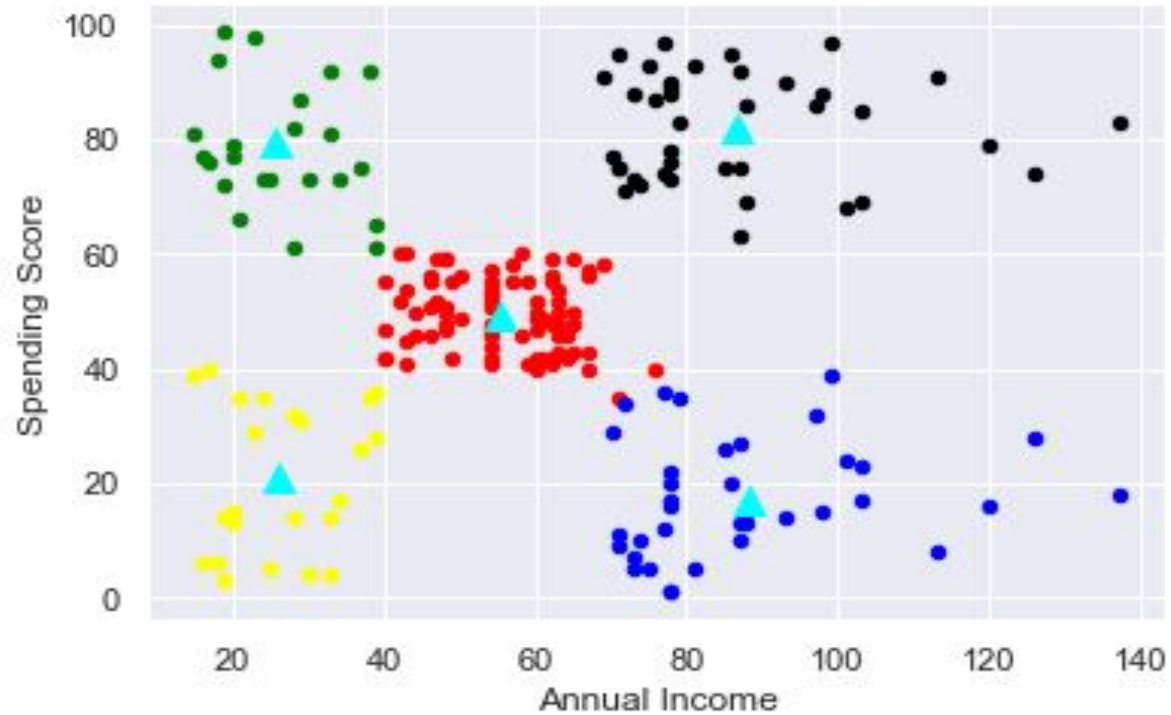
Step-3: Assign each data point to their closest centroid, which will form the predefined K clusters.

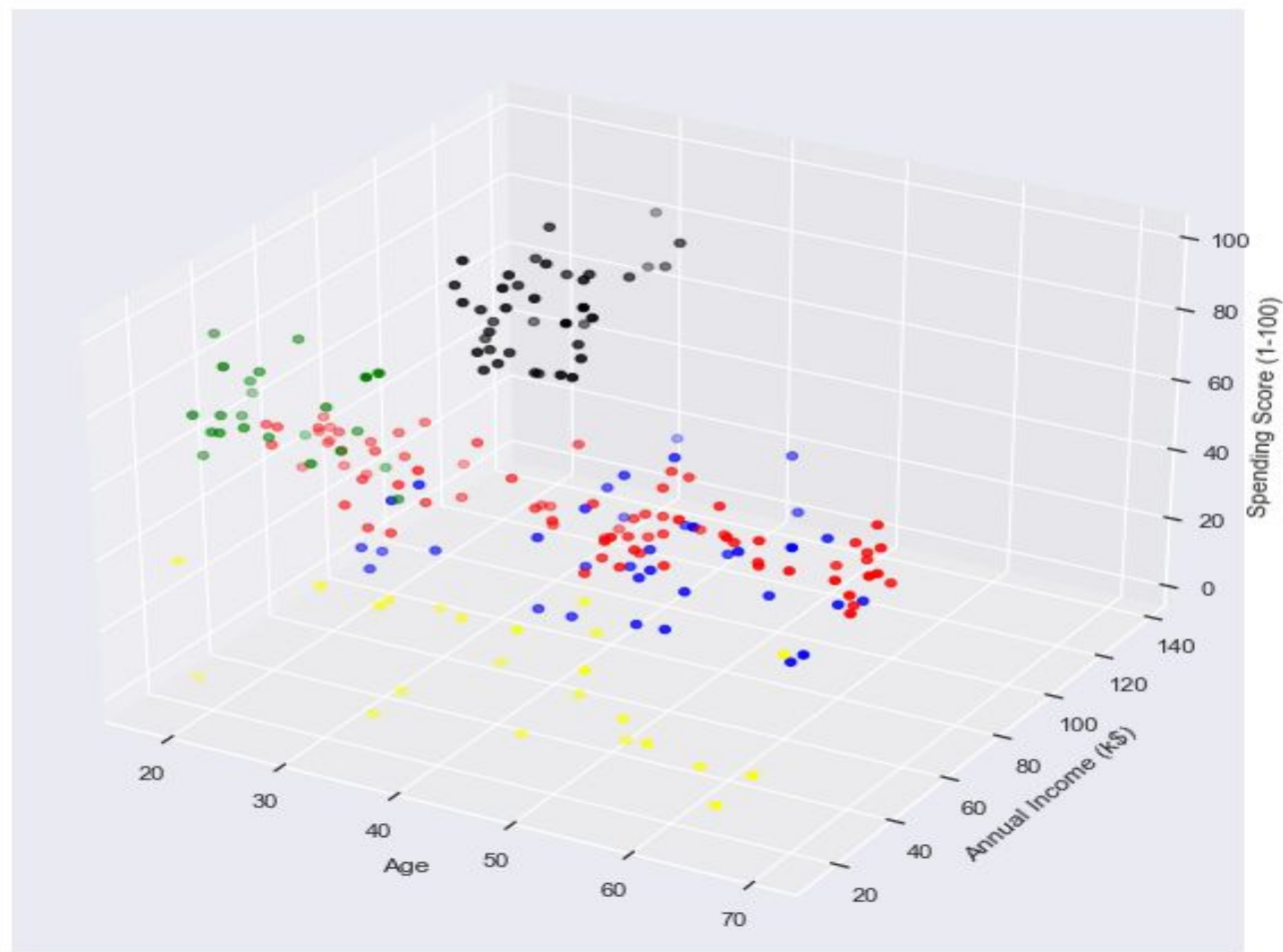
Step-4: Calculate the variance and place a new centroid of each cluster.

Step-5: The model is ready.



RESULT





CONCLUSION

K means clustering is one of the most popular clustering algorithms and usually the first thing practitioners apply when solving clustering tasks to get an idea of the structure of the dataset. The goal of K means is to group data points into distinct non-overlapping subgroups. One of the major application of K means clustering is segmentation of customers to get a better understanding of them which in turn could be used to increase the revenue of the company.

