

CLIENT SEGMENTATION USING MACHINE LEARNING

Dr. M. Saraswathi ¹ Mr. P. Santhosh ² Mr. V . Shyam Sundar³

Assistant Professor, Department of Computer Science & Engineering,
B.E. Students, Department of Computer Science & Engineering
Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya University,
Enathur, Kanchipuram, Tamilnadu

ABSTRACT

The process of grouping clients into various segments who share common characteristics is called Client Segmentation. This enables to create targeted marketing for a specific group of clients which will help in increasing the chances of the person buying a particular product. It allows them to create a specified channel for various segmentation and become more user friendly. An easy example would be that the companies will try to attract the young people a lot through social media posts and older people maybe on radio advertising. This will help in establishing better relationship with the client and their overall performance of an organization. Client Segmentation is such strategy which helps in identifying groups of similar consumers based on their interactions with the product and then effectively implementing various marketing strategies for the suitable people.

Keywords – Clustering , K Means , Segmentation , Visualization

1. INTRODUCTION

Client Segmentation means grouping the clients based on marketing groups which shares the similarity among clients. To be more exact, it means segmenting clients sharing the common characteristics which is the best way of marketing. Client segmentation is gathering information about each client and analyzing it to identify the different patterns for creating the segments. The basic information which includes billing information, shipping information, products purchased, promo codes, payment method etc., Clients are divided into groups based on various characteristics.

We used machine learning algorithm for this client segmentation. One of the most significant uses of unsupervised learning is client segmentation. Companies may discover different categories of clients using clustering algorithms, enabling them to target the prospective user base. We utilized K- means clustering in this machine learning project, which is the most important approach for grouping datasets.

1.1 SCOPE OF THE PROJECT

Grouping of customers is done on the basis of their characteristics, such as geographic, demographic, habits, income, behaviour. To identify various preferences and the first priority of customers. Determining marketing strategies to target specific group of people based on their interests. To make the activities more user friendly that is customer oriented. Customer orientation Makes marketing segment a very important pillar of the Marketing Analogy concept. Identify the areas where the new customers may be created and the market can be expanded.

2. EXISTING SYSTEM

The problem statement will demonstrate the necessary issue for a company operating in an Big industrial market of becoming more client focused. It does not establish the right data collection and analysis process making this segmentation One-Dimensional. It has too many Segmentation Parameters (i.e.) Unwanted criteria or systems are provided for the fitting algorithm which reduces its efficiency. It lacks the differentiation between each cluster. It does not align the Business Objectives with Segmentation Objectives (i.e.) It does not satisfy the business objective to target the specific group of clients.

3. PROPOSED SYSTEM

Machine learning methodologies are a great tool for analyzing client data and finding insights and patterns. Artificially intelligent model is one of the powerful tool for decision-making. They can precisely identify client segments, which is much harder to do manually or with conventional analytical methods. There are many machine learning algorithms, each suitable for a specific type of problem.

One very common machine learning algorithm that's suitable for client segmentation problems is the k-means clustering algorithm. The advantage here is whole process will be automated, so human error will be avoided. Cluster detection may aid in the development of a customized strategy for each cluster base. By tracking clients across months and recognizing the number of clients migrating from one cluster to the next, clustering may also be used to determine client buying behavior.

3.1 System Architecture

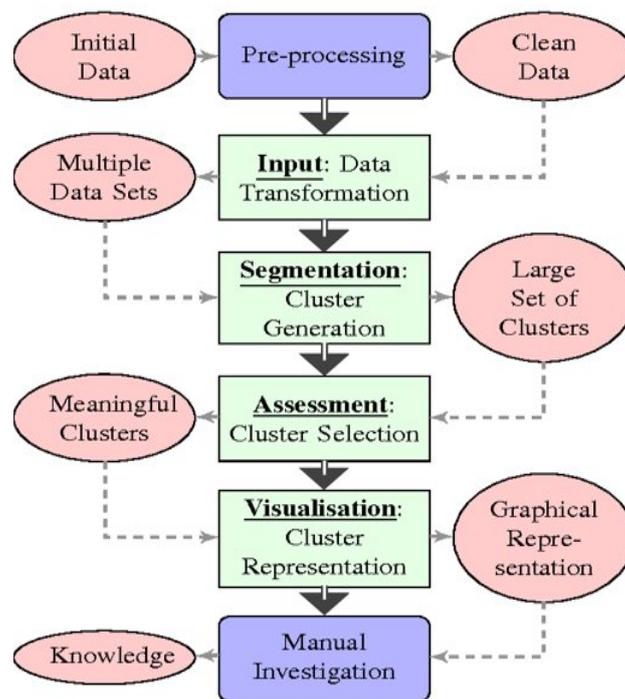


Fig 1. System Architecture

3.2 ALGORITHM

K-means clustering algorithm

The K-means clustering algorithm clusters the given data by separating data points in k groups of equal variances. This effectively minimizes a parameter named inertia. Inertia is within-cluster sum-of-squares distances in this our case. we use the elbow method to find number of clusters as the number of clusters increases until a significant drop in the values of inertia. The stage at this number of clusters is called the elbow of the clustering model. For implementing the elbow method, the below function named “try_different_clusters” is created first. It takes two values as input data K (no.of clusters), data (input data).

Elbow method

The elbow method is used to find the value of the optimal number of clusters. This represents how the generated clusters are spread from one another. The K-means algorithm is evaluated for multiple values of k , and within-cluster sum of square values is calculated for each value of k . After this we will proceed to plot the K versus the sum of the square values. After analyzing this graph, the number of clusters which we need is selected, so that addition of a new cluster doesn't change the value of the sum of square value significantly. Average silhouette method is a approximate measure of how easily each data point fits its corresponding cluster.

SILHOUTTE METHOD

This Method is a measure of how exactly each data point fits its corresponding cluster and also evaluates the quality of clustering. This method evaluates the quality of clustering. Higher the average of silhouette width means better clustering output.

3.3 MODULE DESCRIPTION

1.Importing Data

The data we are going to use features as the Client_seg_data file. It contains various combinations of values corresponding to the respective entity. For this project, our primary packages are going to be Pandas to work with data, NumPy to work with arrays, scikit-learn for data splitting,

Seaborn works easily with data frames and the Pandas library. The graphs created can also be customized using the seaborn and finally, the algorithm modules are imported by a special package called yellowbrick . Import all the necessary packages into our python environment.

2. Feature Selection & Data Split

We need a data manipulation plan to implement clustering on our data. We'll use our hypothesis to guide us. We need to get the data frame into a clustering format and then we'll manipulate into other variables.

3. Modeling

First, we will randomly initialize the value of k as the number of clusters . After we allot each data points to the closest centroid forming separate groups while relocating the center to the middle of all cluster employing euclidian distance. While working through the preceding steps, the algorithm will try to reduce the sum of squared distances among clustered-point and middle for all clusters. When all data points unite the repetition will end.

4. Evaluating the created clusters

Cluster analysis is a process of using a mathematical model to discover groups of similar data based on finding the smallest variations among datas within each group. Scaling the data, similarity measure, Visualization of Pair-wise Distances and Robustness analysis are done as a part of Cluster evaluation.

4. RESULT

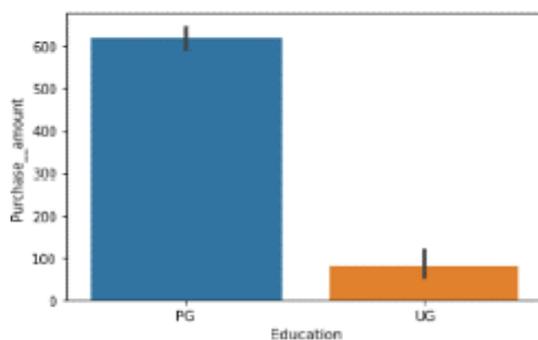


Fig 2. Purchase amount and Education

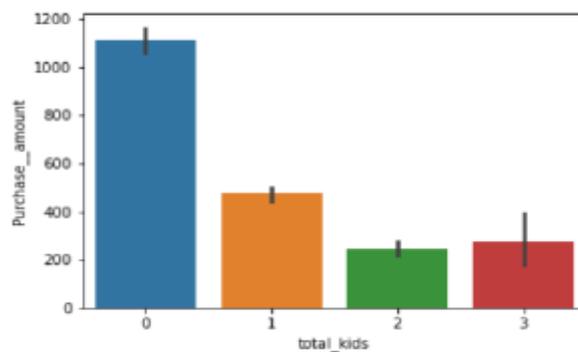


Fig 3. Purchase amount and Total Kids



Fig 4. Heat Map

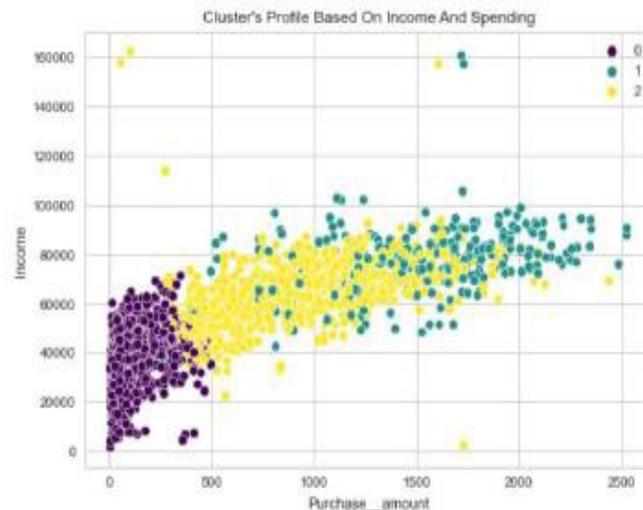


Fig 5. Cluster Profile based on income and spending

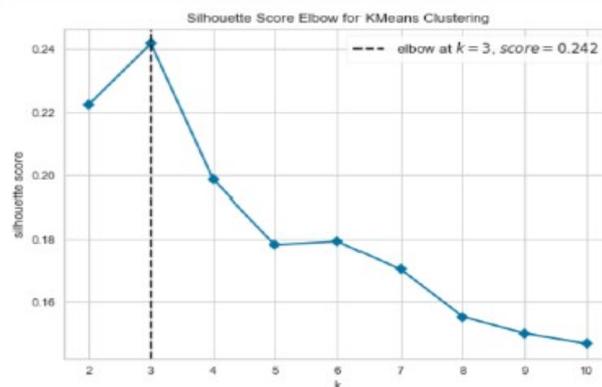


Fig 6. Silhouette Score elbow for KMeans Clustering

5. CONCLUSION

In this Paper, we went through the client segmentation model. We developed this using Machine learning. Specifically, we made use of a clustering algorithm called K-means clustering. We analyzed and visualized the data and then proceeded to implement our algorithm. Our model has partitioned client data into mutually exclusive groups, three clusters in our case. This will be useful for applying further data mining strategies and the derived insights are helpful in decision making for the business wings.

6. REFERENCES

1. Fuzzy K-Means clustering algorithm for data with unequal cluster sizes and contaminated with noise and outliers: Review and development Salar Askari Expert Systems with Applications 165, 113856, 2021
2. T. Kansal, S. Bahuguna, V. Singh and T.Choudhury, "Customer Segmentation using K -means Clustering," International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS), Belgaum, India, 2018
3. Visualization of education blockchain data: trends and challenges Patrick Ocheja, Brendan Flanagan, Hiroaki Ogata, Solomon Sunday Oyelere interactive Learning Environments, 1-25, 2022
4. Machine learning based classification and segmentation techniques for CRM: a customer analytics Narendra Singh, Pushpa Singh, Krishna Kant Singh, Akansha Singh International Journal of Business Forecasting and Marketing Intelligence 6 (2), 99-117, 2020
5. Implementation of K-means algorithm for clustering corn planting feasibility area in south lampung regency by AA Aldino, D Darwis, AT Prastowo, C Sujana Journal : Conference Series 1751 (1), 2021
6. Selecting optimal k in the k-means clustering algorithm Mojtaba Jahanian, Abbas Karimi, Faraneh Zarafshan ,Journal of Computer & Robotics, 31-42, 2022
7. An entropy-based initialization method of K-means clustering on the optimal number of clusters Kuntal Chowdhury, Debasis Chaudhuri, Arup Kumar Pal Neural Computing and Applications 33 (12), 6965-6982, 2021
8. Clustering in PySpark by Pramod Singh in Machine Learning with PySpark, 127-155, 2022
9. K-means algorithm for clustering of learners performance levels using machine learning techniques. Revue d'Intelligence Artificielle, Vankayalapati, R., Ghutugade, K.B., Vannapuram, R., Prasanna, B.P.S. (2021). Vol. 35, No. 1, pp. 99-104.

Author's Profile:-



Dr. M. Saraswathi Assistant Professor in Computer Science and Engineering
Sri Chandrasekharendra SaraswathiViswa Mahavidyalaya deemed to be university,
Enathur, Kanchipuram, India.



Mr. P. Santhosh , Student, B.E. Computer Science and Engineering,
Sri Chandrasekharendra SaraswathiViswa Mahavidyalaya deemed to be university,
Enathur, Kanchipuram, India.



Mr. V. Shyam Sundar , Student, B.E. Computer Science and Engineering, Sri
Chandrasekharendra SaraswathiViswa Mahavidyalaya deemed to be university,
Enathur, Kanchipuram, India.