



BEHAVIOURAL CUSTOMER SEGMENTATION BASED ON MACHINE LEARNING ALGORITHM

Gandhodi Harsha Vardhan^a, Morthala Kala Rushitha Reddy^a, Maddali Surya Teja^a,

Sadda Sai Mokshitha^a, Ponnada Pavani^a,

Department of Computer Science and Engineering (Data Science),

Jani Deemed-To-Be-University, Bangalore, India.

Abstract: Nowadays the customer segmentation became a very popular method. The method of breaking up consumers into smaller groups is known as customer segmentation, often referred to as market segmentation. This makes it simpler to tailor your messaging to the demands of your audience. Behavioral customer segmentation is the process of separating customers into groups based on any of the following qualities is known as behavioral segmentation. {Utilization, fidelity, mindfulness, occasions, education, and purchase intentions}. Customers are identified using a computer technique called the k-means clustering algorithm based on their behavioral characteristics. Developed clusters assist the business in identifying certain clients and selling its content to them through social websites with media in which they are genuinely interested. different clustering algorithms (k-Means, Agglomerative, and Mean shift) are been implemented to segment the customers and finally compare the results of clusters obtained from the algorithms.

Index Terms – Customer Segmentation, Machine Learning, K-Means Algorithm, Agglomerative.

INTRODUCTION

Today, a lot of businesses operate online, making internet marketing crucial for retaining clients. However, it is not very effective to treat all customers equally and use the same marketing tactics to target them all. Instead, it irritates the customers by disregarding their uniqueness, hence customer segmentation is quickly gaining popularity and emerging as an effective remedy for this issue. Customer segmentation is the process of categorising a company's clientele based on behavioural (such as the kinds of items they order and their yearly income) and demographic (such as age, gender, and marital status) characteristics. Behavioural factors are a superior technique for customer segmentation since demographic indicators do not place an emphasis on the individuality of the consumer because same-age groups may have distinct interests.

K-Means Clustering:

It is the most basic clustering technique based on the partitioning concept. The algorithm is affected by Initialization of the centroids position, the number of K (centroids) is calculated by the elbow method (discussed further in this section), after calculation of K centroids in terms of Euclidean distance data points are assigned to the closest centroid forming the cluster, after the cluster formation the barycentre's are calculated by the cluster and this process is repeated until there is no change in centroid position.

LITERATURE SURVEY AND REVIEW

The business world has been more competitive over time since companies like these need to satisfy the wants and demands of their existing clients in order to grow their customer bases. It may be quite challenging to recognise and satisfy each customer's wants and objectives. This is due to the fact that different clients have different demands, goals, demographics, sizes, preferences, and other characteristics. Currently, it is not a good business strategy to serve every consumer equally. The notion of market or customer segmentation, in which customers are broken down into smaller groups or segments and the members of each

subcategory have common market traits or behaviours, has been embraced for this difficulty. As a result, the process of splitting the consumer base into indigenous groups. For example, Mall Customer data is an interesting dataset that has hypothetical customer data. It puts you in the shoes of the owner of a supermarket. You have customer data, and on this basis of the data, you have to divide the customers into various groups. therefore, the data includes the following features:

1. Customer ID
2. Customer Gender
3. Customer Age
4. Annual Income of the Customer
5. Spending score of the Customer

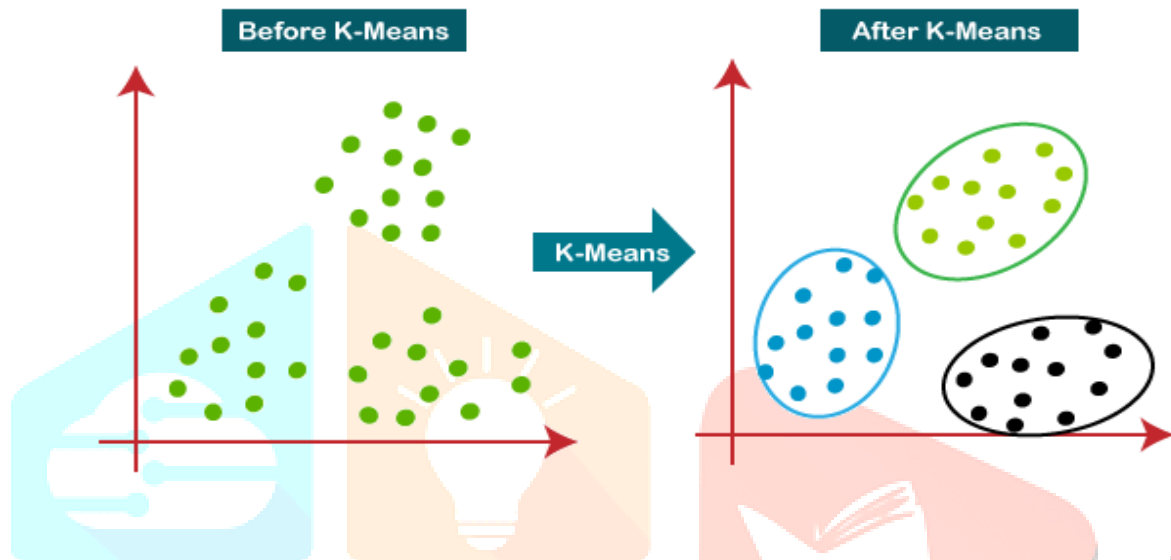
Review:

- **Customer segmentation:**
- Harsha et al. explains that customer segmentation involves dividing consumers into groups based on marketing similarities among members of the same group. Based on a variety of factors, including age, hobbies, purchasing patterns, and other factors, customers are grouped into distinct clusters.
- **Rushitha et al lists the four widely used segmentation bases:**
- 1) Geographic Segmentation: dividing a population into groups according to climate, population density, or geographic area.
- 2) Demographic Segmentation: Determine market segments based on factors such as age, size, and family structure.
- 3) Psychographic Segmentation: Customer life style factors, such as interests, views, attitudes, etc. are used to segment customers.
- 4) Behavioural Segmentation: This kind of segmentation bases its conclusions on real consumer behaviour toward a product, such as brand loyalty, user status, purchase readiness, etc.
- By utilising the benefit of segmentation, marketers may benefit from the divide and conquer approach, which is the foundation for customer segmentation. edge over a specific market group and gradually beat out rival marketers. Market segmentation enables marketers to concentrate more on client relationship management, something that was previously impossible with current mass marketing strategies.
- **Clustering:**
- Mokshitha et al. Clustering is the act of putting information in a dataset into groups based on certain similarities. Depending on the circumstances, one may select from a variety of methods to use on a dataset. However, there is no one

clustering method, therefore choosing the right strategies for clustering becomes crucial. and others Utilizing the Python Scikit-Learn module, we constructed three clustering methods in this study.

- **The Clustering Explained:**

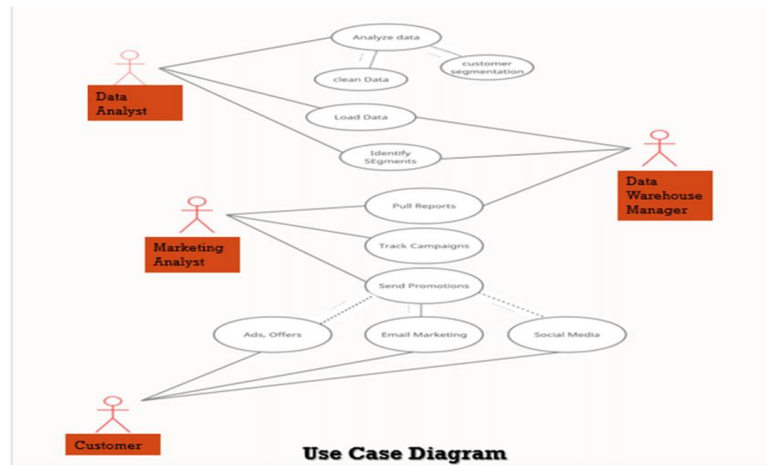
The various aspects of how the algorithms to cluster data can be tuned and modified, and clustering algorithms try to find natural clusters in data. The foundation of clustering is the idea that objects in the same cluster must be related to one another. Related elements are placed close to one another in the data grouping.



Uses of Clustering:

- **Marketing:** With the help of existing customer data, clustering can be used in the marketing industry to identify different customer groups. Customers may receive discounts, offers, promo codes, and other things based on that.
- **Real Estate:** Using clustering, it is possible to categorise and rank different property locations according to their worth and significance. Clustering algorithms can go through the data and isolate different groupings of real estate based on their prices.
- **Book Store and Library Management:** Clustering can help libraries and bookshops manage their book databases more effectively. Better operations can be implemented with the right book ordering.
- **Document Analysis:** We frequently must classify different study papers and materials according to similarities. And we lack terminology in such circumstances. Additionally impossible is the manual labelling of enormous volumes of data. The programme may analyse the text and divide it into groups by using clustering into different themes.

Use Case Diagram:



Use case diagram of proposed system consist of 4 users 1. Data Analyst 2. Marketing Analyst 3. Data Warehouse Manager 4. Customer in figure 1 And 6 use cases.

1. Analyse Data: Analyst has the access to loaded data and analyst clean the data and perform analysis to form clusters.

2. Load Data: Analyst log into database & view data & load into memory to work on it.

3. Identify Segments: Analyst form report for segmented customer data and send to data warehouse and marketing analyst can access that data to form marketing strategies.

4. Pull Reports: Marketing team can view & make edits on the reports, data for report is pulled from DW system. Customer Segmentation Using Machine Learning.

5. Track Campaigns: The customer's interaction tracked by marketing team for success report.

6. Send Promotions: Marketing team send promotions through mail, social media ads, paid ads, coupons.

METHODOLOGY:

Data Collection: The dataset has been taken from a local retail shop consisting of two features, average number of visits to the shop and average amount of shopping done on yearly basis.

Feature scaling: The data has been scaled using Standard Scaler by applying standard scaler the data gets centred around 0 with standard deviation of 1.

X-Means(x)/Stdev(x)

x =Entry in a feature set x_i of X

$\text{Mean}(X)$ =Mean of feature set X

$\text{Stdev}(X)$ =Standard deviation of X

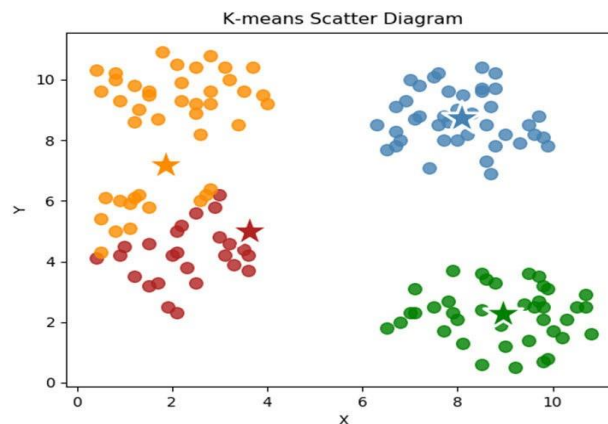
K-Means Clustering Algorithm:

K-means Clustering is a clustering Algorithm in which we are given with data points with its data set and features and the mechanism is to categories those data points into clusters as per their similarities. The algorithm forms K Clusters based on its similarity. To calculate the similarity K-Means uses Euclidean distance measurement method.

1. In first step, we randomly initialize k points.
2. K-means classifier categorizes each data point to its nearest mean and rewrite the mean's coordinates.
3. Iteration is continuing up till all data points are classified.

Agglomerative Clustering:

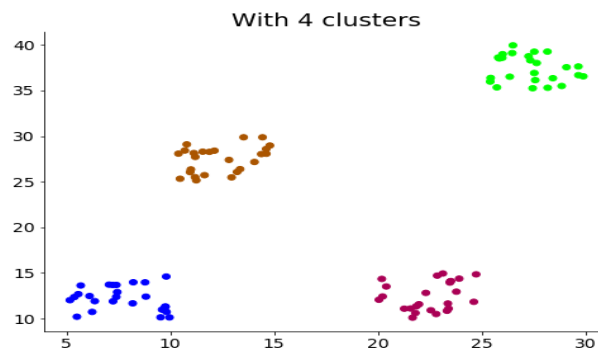
- Choosing the optimal number of Clusters: Cluster value for this algorithm have been calculated by the Dendrogram as described in Dendrogram section which also gave the value of $K = 4$



Algorithm:

1. Each data point is taken as to be a cluster.
2. Merge the two closest cluster.
3. Needs to be repeated until all the data points are merged together to form a single cluster.
4. However, as we have defined the value of K as 4, the algorithm will stop when all the data points are part of any of the 4 clusters.

Example Diagram:



Since the number of clusters for K-means and agglomerative models is equal, they produced final clusters with the same pattern. But if you pay close attention, a point. The top left corner of the Standard cluster has a new cluster, and the lower right corner of the Standard cluster has a similar situation with 2–3 points.

- K-Means Clustering:

Surya Teja et al. One of the most well-liked partitioning clustering algorithms is the K-Means method. This grouping Each data point is placed on one of the K non-overlapping clusters that are chosen before the algorithm runs, and the clusters that are generated correlate to the hidden pattern in the data that provides the necessary information to aid in decision-making.

- Agglomerative Clustering:

This clustering falls under the category of hierarchical clusters, which are created using a hierarchy. Hierarchical clustering is based on the idea that objects close to one another are more related to one another than objects far from one another. Harsha et al. The fundamental problem with the hierarchical technique is that once a split or merge operation is performed, it cannot be undone. This challenge is advantageous since it results in lower processing costs by avoiding the need to consider a combinatorial number of options. Surya et al. Top-down approach, also known as divisive clustering, and bottom-up strategy, also known as agglomerative clustering, are the two strategies used in hierarchical clustering.

- Mean shift Clustering:

Pavani et al. specifies the gradient ascent method used in this approach. In mean shift, discrete data samples are used to find the local maxima of a density function. It operates by placing a search window over a portion of the distribution. The mean shift approach is employed for actual data analysis, which is a tool that is application dependent and does not initially presume the structure of the data cluster. Wide-ranging applications of this technique include object identification and picture segmentation.

Conclusion:

Customer segmentation is performed on the company's customer data, and customers are divided using features such as total spending and annual income. This study also demonstrates that dividing customers based on behavioural characteristics is a better solution for the existing customer segmentation problem, and the K-means clustering machine learning algorithm is identified as a good choice for this approach. Internal cluster validation may be used to select the optimal clustering technique for the dataset and to accurately cluster data into its opposite cluster.

Acknowledgment

I would like to express my special gratitude to our guide Dr. Raja Praveen Kumar sir as well as our Project Centric Learning coordinator Dr. Vijay Kumar sir who gave me the golden chance to do this magnificent project on the topic BEHAVIOURAL CUSTOMER SEGMENTATION BASED ON MACHINE LEARNING ALGORITHM, which also helped us in doing a lot of Research and we came to know about so diversified things. We are so pleased to them.

Furthermore, we worked hard for the accurate output.

REFERENCES

- [1] Sukru Ozan, "A Case Study on Customer Segmentation by using Machine Learning Methods", IEEE, Year: 2018.
- [2] Jayant Tikmani, Sudhanshu Tiwari, Sujata Khedkar "Telecom customer segmentation based on cluster analysis An Approach to Customer Classification using k-means", IJIRCCE, Year: 2015.
- [3] Domavicius, G., and Tuzhilin, A. (2015). Context-aware recommender systems. In Recommender systems handbook (pp. 191-226). Springer US.
- [4] Hwang, Y. H. (2019). Hands-on Advertising Science Data: Develop your machine learning marketing strategies... using python and r. S.I: Packt printing is limited.
- [5] Vaishali R. Patel and Rupa G. Mehta "Impact of Outlier Removal and Normalization Approach in Modified k-Means Clustering Algorithm", IJCSI, Year: 2011.
- [6] Omar Kettani, Faycal Ramdani, Benaissa Tadili "An Agglomerative Clustering Method for Large Data Sets", IJCA, Year: 2014.
- [7] Rivedi, A., Rai, P., DuVall, S. L., and Daume III, H. (2010, October). Exploiting tag and word correlations for improved webpage clustering in Proceedings of the 2nd international workshop on Search and mining user-generated contents (pp. 3-12). ACM.
- [8] H. Mehta, V.S. Dixit and P. Bedi, "Refinement of recommendations based on user preferences".
- [9] Yogita Rani and Dr. Harish Rohil "A Study of Hierarchical Clustering Algorithm", IJICT, Year: 2013.
- [10] Blanchard, Tommy. Bhatnagar, Pranshu. Behera, Trash. (2019). Marketing Analytics Scientific Data: Achieve your marketing objectives with Python's data analytics capabilities. S.I: Packt printing is limited