

# RESEARCH PAPER ON CUSTOMER SEGMENTATION

## AIM

The aim for this research paper is to “develop an extensive and through understanding on customer behaviour or segmentation by examining various parameters which will helps in making an organisation grow and lets the old organisation to stay in trend by analysing customer behaviour and change in their needs. This helps an organisation to stay ahead from the rest. This will investigate the influence of psychographic factors on consumer purchasing decisions and to develop a segmentation model that incorporates these factors to enhance marketing strategies."

## OBJECTIVE

There are few objectives for this research paper as given under :-

- I. To identify and analyse the key demographic, psychographic, behavioural, and geographic variables that influence customer segmentation in the retail industry.
- II. To identify the difference in parent and non – parent’s requirements and show that how parents with children of different age groups makes their choices and affects their choice in products.
- III. It also reflects the people which come under the different incomes such as low income, high income affects the segmentation.

## ABSTRACT

In today’s era where innovation is keeps on occurring and new products are innovated to fulfil the customers needs, there are various product present for a particular requirement which makes it tough for already existing company to remain in the game. To grow more and earn profit an organization need to know the portion they need to work on and attract a good number of new audience while satisfying the existing one. To make this happen, Machine Learning plays an important role by unrevealing the hidden insights and trends to figure the fields where they are weak and where they want to bring new creative idea too. Organization has to significant challenges as new start – ups are knocking the markets and facing on every micro requirement which makes the market very competitive.

For survival, company need to know their targeted audience and what product to sell for that specific group of audience. For the accomplishment of this goal machine learning Models are great help and for this paper ,clustering is used to group the data with similar interest in one cluster and let the insight make some sense. A large dataset of around 2240 is being used to get the conclusion out of it.

It consists of many parameters and necessary information which makes it significant to delve into and by applying clustering via Agglomerative clustering which comes under Data Mining where clusters are kept on combing until all comes together. With this a conclusion will come out based on the 4 profiling which will be consisting cluster number 0,1,2 and 3 . Each cluster number has different insight assigned to it which are their family structure and type of income.

## KEYWORD

Machine Learning, Hierarchical clustering algorithm, Unsupervised algorithm, Agglomerative clustering

## INTRODUCTION

Customer Segmentation is simply defined as grouping people or customers based on their income, age, family, etc., to get an idea about their buying or requirements. As the competition is growing and more complications coming in a way of companies it becomes necessary to have a thorough understanding about the customers choices and their requirements, which helps company to grow and stay in competition. Customer Segmentation gives a thorough understanding about the places the need to work on and deals with any ups and downs. Here, a cluster is formed with the similar taste and like this one can easily focus on any set of customers, on to their requirement and changing requirement as well. In this research paper, there are few objectives which initially focuses on to the segmentation of customers based on their similarity or patterns which makes it easier to cluster them. It also reflects that how their lifestyle change can affect their buying by deepening the understanding about their psychological behaviour. It will also reflect that how machine learning can play a major role in order to bring out some great insights which the company needs to follow and enjoy great profits.

For making it sound and successful it is important to take corrective measures which leads to various heights such as depending on to the unsupervised learning, which is known to deal with unlabelled data and produce output out of it. Some of the Machine Learning Algorithm is also used which includes K-means clustering and Agglomerative Hierarchical clustering. Both algorithms are crucial in themselves and hence make application of it in problem statement like customer segmentation and more. Agglomerative Hierarchical cluster works on Bottom – up approach which means it starts with atomic clusters and soon it turns out to be a big cluster by combining similar clusters together that contains all the data and parameters. It plots dendrograms which are tree-like structures that stores similar clusters together and begins from single clusters as the similarity is being found more clusters are combined.

Clustering is best known to deal with unlabelled data where clusters are formed based on the similarity which makes it easier to work on and hence used to solve various real life problems which includes Customer Segmentation which helps the company to have an idea about what their old customers are keeps on buying repeatedly and what changes they asked as per the change in their life style and various parameters which include their age or from being married to parents of child. It will let the organisation to have an overview that as the circumstances are changing so as the requirements. Which makes an organisation ready to face any challenge and become resilient, it helps them to stay strong in the competition and works great under all the changing circumstances. It lets the company to have an idea what are their target audience is and how to bring the customer satisfaction on the table and to brainstorm on their current marketing tactics and skills to bring out the best which serve them with fruitful results. It also allows them to figure out the influencing factor or X factor which plays a crucial to increase the sales or affect the sales, and as soon they find out this protects them from facing any inconvenience in the future.

## NOVELTY STUDY

For this research paper, introducing a new machine learning algorithm of Agglomerative Hierarchical clustering for customer segmentation that outperforms existing methods in terms of accuracy and efficiency.

[1] A solution is proposed as distinguish the customer group into two groups named as premium and standard with the help of machine learning methods named as NEM and LiEM. In this paper a customer segmentation is achieved from various parameters such as family status and different age group of children.[3] Deftly positioned within the domain of unsupervised learning, clustering emerges as a potent technique for unravelling latent structures within data. Dispensing with labelled

responses, unsupervised learning methods strive to discern underlying patterns and groupings that permeate a dataset.

## METHODOLOGY

For accomplishing this task, we need some technical skills and algorithm to apply. Following is the detailed overview on all those methods which are used in this research paper:-

### 1. Python

Python is known for its versatile nature and ease to use features. Python is a high-level programming language which offers easy syntax which various other features makes it in demand and is also beginner friendly. Python is used to work with machine learning algorithm which we used here as well. Due its large repository and libraries which it provides makes it significant to perform machine learning program using it.

### 2. Machine learning

Machine learning is simply a process to learn from past data and make an output out of it. It does not require any extensive coding or programming to make a model work instead it trains the model from past data and let the machine learn out of it. [4] Rather than relying on explicit programming instructions, machine learning systems leverage patterns and statistical models to make predictions or decisions. This adaptability is a key feature that distinguishes machine learning from traditional rule-based programming. Machine Learning is known for dealing with labelled and unlabelled data by using various types of machine learning algorithm.

There are various types of machine learning algorithm that consist different working procedure and skills which makes aside from each other.

Following is those machine learning algorithm :-

#### i. Supervised Learning

Supervised Learning is a type of machine learning which deal with label data and make the prediction out of it. It only deals with labelled data where by getting an input (X) it provides an output(y). It simply gives the output in categorical or numerical value such as 0,1 etc.

#### ii. Unsupervised Learning

Unsupervised Learning is known to deal with unlabelled data. It read unlabelled data and trains the model by gathering or collecting the data with similar features together and forms a cluster out of it. Unsupervised learning does not require any guidance or instructions from humans as it find out the insights on its own from clusters and associations.

### 3. Hierarchical Clustering

Hierarchical clustering is a type of unsupervised machine learning which deal with unlabelled data where it forms clusters and based on it, gives an output. It deals to give insights from the clusters itself and makes a hierarchy of data as it follows Bottom-up approach – it begins from single clusters of data and as it finds some similarity it keeps on making clusters and association out of it.[4] Hierarchical clustering creates a tree-like structure of clusters, known as a dendrogram, by iteratively merging or splitting clusters. It does not require specifying the number of clusters in advance and offers a hierarchical view of data grouping.

Hierarchical Cluster follows two approaches :-

#### i. Agglomerative approach

It is a type of Hierarchical clustering technique which follows bottom-up approach where the dataset from being in its atomic format combines and create a hierarchy out of it. It is easier to make cluster in this algorithm as

there no need to predetermine the number of cluster or no need to makes all the clusters of same size which are few cons of K-means clustering.

Following is the algorithm which is followed by agglomerative approach :-

- a. Create single or atomic cluster from the given data points. As if we have N datapoints so there will be N clusters.
- b. Now, look for cluster with have similar features or which are not far from each other as one. Now your cluster will become N-1 in number.
- c. Again, perform the same step as (b) and merge the two closest clusters together and it will become N-2 in numbers.
- d. Keep repeating the same step as earlier (b) till there will only one cluster left.
- e. Finally, when all the cluster will be formed into one large cluster, begins with the creation of dendrogram and make hierchary of cluster out of it.

ii. Divisive Algorithm

This is the second approach followed under hierarchical cluster to deal with unlabelled data. It has different approach from agglomerative approach as it follows top-down approach. It begins from a large cluster and keeps on splitting it into smaller clusters. It is very efficient and accurate method to deal with clusters.

## PROPOSED SYSTEM

### I. Data Collection and Preparation

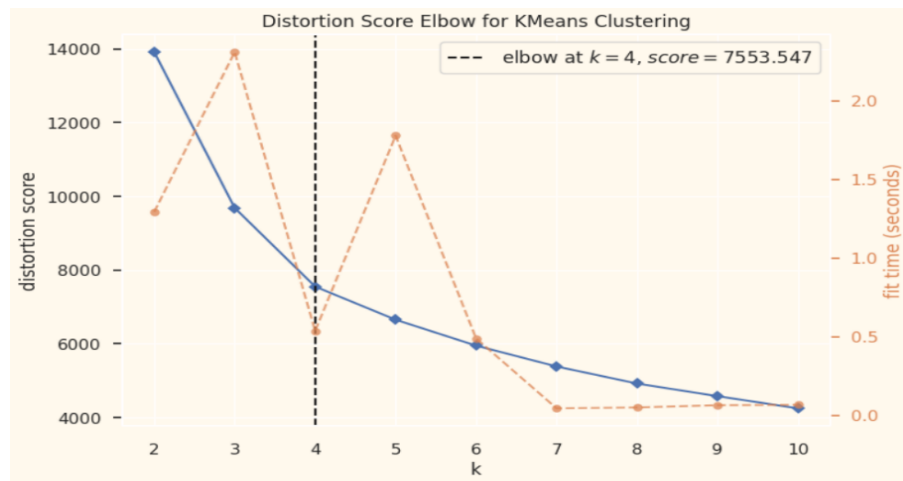
This is the first step to begin train a machine learning model as here data cleaning and preparing will be done. Data cleaning is done to remove all the ambiguities, NA value or missing value from the given dataset. Ones our data become clean the further process will become even more easier. Secondly, Data preparation was

- Label encoding the categorical features
- Scaling the features using the standard scaler
- Creating a subset data frame for dimensionality reduction takes place.

### II. Feature Extraction

This is very crucial step to make our model more accurate and make it efficient. As it chooses or extract information from the existing data points and brings out accuracy out of it. Choosing the correct features out of it is necessary in our case we have extract age, spends, living with, children, family size and parenthood out of the given data.[3] . Feature engineering empowers you to de rive insights from the data that might not be immediately apparent, ultimately enhancing the model's ability to capture complex relationships and patterns.

### III. Clustering



- Elbow Method to determine the number of clusters to be formed [2] during the formation of groups it is mandatory to select optimal no of clusters we applied hyper tuning which is achieved by Elbow Method to choose optimal no of clusters.

In the given graph the elbow method shows a bend in graph at 4 and from their it becomes flat which shows *that 4 is optimal no of cluster.*

- Clustering via Agglomerative Clustering
- Examining the clusters formed via scatter plot

### IV. Data Visualization

Lastly, we will represent our given datapoints in the form of visual representation by using various graphs and plots, for this we have used dendrograms, bar graphs, scatter plot, swarm plot, count plot etc.

## CONCLUSION

In this research paper we have concluded that the whole data is divided into four clusters where each has its unique features. If we talk about **cluster 0** so it contains customers which are parent and older with 4 members in their family and most of them have teenagers. **Cluster 1** contains a set of customers which are not parent and have at most 2 members in their family majority of them are partners mostly of all ages of people stays in this group with high income. **Cluster 2** contains parents with a family of three and have one child at most. **Cluster 3** are parents with 5 members in their family consist of teenagers with low income and old in age.

## REFERENCE

- [1] Sukru Ozan, "A Case study on Customer Segmentation by Using Machine Learning Methods", IEEE, Year, 2018
- [2] Patankar, Nikhil, et al. "Customer Segmentation Using Machine Learning." Recent Trends in Intensive Computing, December 2021, DOI:10.3233/APC210200. Licensed under CC BY-NC 4.0, ResearchGate.
- [3] Korn, Jonathan Wayne, PhD. Machine Learning in Python for Everyone. Amazon Digital Services LLC - Kdp, 2023. ISBN 9798869937872. 122 pages. Computers, Machine Theory.

[4] Baah, Bernard. *Python Machine Learning Essentials*. Volume 1 of Programming, Data Analysis, and Machine Learning, Amazon Digital Services LLC - Kdp, 2024. ISBN 9798884780767.