# CUSTOMER SEGMENTATION USING UNSUPERVISED MACHINE LEARNING IN PYTHON

[1] Dr.S.Kavitha M.Sc., M.Phil, Ph.D., Head of the Department,,

[2] P. Kiruthika, M.Sc (Computer Science),

[1] [2] Department Of Computer Science, Sakthi College of Arts and Science for Women, Oddanchatram.

## ABSTRACT

The emergence of many competitors and entrepreneurs has caused a lot of tension among competing businesses to find new buyers and keep the old ones. As a result of the predecessor, the need for exceptional customer service becomes appropriate regardless of the size of the business.[2] Furthermore, the ability of any business to understand the needs of each of its customers will provide greater customer support in providing targeted customer services and developing customized customer service plans. This understanding is possible through structured customer service. Each segment has customers who share the same market features.[5] Big data ideas and machine learning have promoted greater acceptance of automated customer segmentation approaches in Favor of traditional market analytics that often do not work when the customer base is very large. In this paper, the k-means clustering algorithm is used for this purpose.[8] The Sklearn library was developed for the k-Means algorithm (found in the Appendix) and the program is trained using a 100-pattern two-factor dataset derived from the retail trade. Characteristics of average number of customer purchases and average number of monthly customers.

## I INTRODUCTION

It is possible to win the competition in the market and enhance corporate earnings by better understanding the client's needs. Companies can develop successful marketing strategies if they are aware of the wants and needs of their target audiences. While the requirements and expectations of each customer are unique, many customers share identical or quite similar qualities. Customer segmentation is one method that may be used to put together multiple different consumers who share similar qualities. Improving the quality of the connections with your customers also requires proper consumer segmentation. Marketing intelligence is conducting information analysis to comprehend better a target market and its consumers'

demographics [1], [2]. In marketing, it is common for analysts to categorize customers into comparable customer groups to understand better how to advertise to each group of customers. Therefore, segmentation is a collection of approaches that might be useful in categorizing different types of customers. Customers' existing relationships with a company are the primary focus of most direct marketing operations. The more you know about your customer's needs, desires, and purchasing habits, the easier it is to tailor marketing programs to their needs and desires and how they buy things. Buried in a database of integrated data proved to be effective for detecting subtle but subtle patterns or relationships. This mode of learning is classified under supervised learning. Integration algorithms include the KMeans algorithm, K-nearest algorithm, sorting map (SOM), and more.[4] These algorithms, without prior knowledge of the data, are able to identify groups in them by repeatedly comparing input patterns, as long as static aptitude in training examples is achieved based on subject matter or process. Each set has data points that have very close similarities but differ greatly from the data points of other groups. Integration has great applications in pattern recognition, image analysis, and bioinformatics and so on.[15] In this paper the k-means clustering algorithm was implemented in the customer segment. The scalar library (Appendix) of the K-Means algorithm was developed, and training was started using a standard silhouette -score with two feature sets of 100 training patterns found in the retail trade. After several indications, four stable intervals or customer segments were identified. Two factors are considered in combination with the number of items a customer purchases per month and the average number of customers per month.

## II LITERATURE SURVAY

Aman Banduni, Prof Ilavedhan A, in [1] studies customer segmentation using machine learning. In this paper, they explained the concept of customer segmentation.

Kamalpreet Bindra, Anuranjan Mishra in [2] studies detailed different clustering algorithms. And compared different algorithms and based on the results decides which algorithms to use for the project.

Kai Peng (Member, IEEE), Victor C. M. Leung, (Fellow, IEEE), and Qinghai Huang in [3] get to know in detail about mini-batch K-means clustering algorithm. Get to know about the advantages and disadvantages of the algorithm and also about the implementation.

Fionn Murtagh and Pedro Contreras in [4] studied hierarchical clustering algorithms. In this paper get to know more about this clustering algorithm and also observe how clusters formed and also about advantages and disadvantages and compare it with the other different clustering algorithms.

D. P. Yash Kushwaha, Deepak Prajapati in [5] studied customer segmentation in detail and also studied in detail about k-means clustering algorithm and performed customer segmentation using K-means clustering algorithm and observed the clusters formed and compared the results with the other clustering algorithms.

Manju Kaushik, Bhawana Mathur in [6] get to know in detail about two different clustering algorithms such as K-means clustering algorithm and hierarchical clustering algorithm. And perform customer segmentation using these two algorithms and compare the results and decide the best clustering algorithm between these two to perform customer segmentation.

Ali Feizollah, Nor Badrul Anuar, Rosli Salleh, and Fairuz Amalina in [7] studied in detail two different clustering algorithms such as K-means clustering algorithm and mini-batch means clustering algorithm. And perform customer segmentation using these two algorithms and compare the results and decide the best clustering algorithm between these two to perform customer segmentation.

Asith Ishantha in [8] studied in detail different clustering algorithms such as K-means clustering algorithm and mini-batch-means clustering algorithm and hierarchical clustering and many more. And perform customer segmentation using all these algorithms and compared the results and decide the best clustering algorithm between all these to perform customer segmentation.

Onur Dogan, Ejder Aycin, Zeki Atil Bulut in [9] studied customer segmentation in detail using the RFM model and some clustering algorithms.

Juni Norma Sari, Ride Dedriana, Lukito Nugroho, Paulus Insap Santosa in [10] reviewed all customer segmentation techniques.

Shi Na; Liu Xumin; Guan Yong in [11] studied in detail about k means clustering algorithm and observed its pros and cons.

Francesco Musumeci; Cristina Rottondi; Avishek Nag et. al in [12] get an overall overview of the application of machine learning techniques and understand their implementation.

Şükrü Ozan et. al in [13] studied about Case Study on Customer Segmentation by using Machine Learning Methods.

Tushar Kansal; Suraj Bahuguna; Vishal Singh; Tanupriya Choudhury in [14] studied mostly customer segmentation using the K-means clustering algorithm.

Ina Maryani; Dwiza Riana; Rachmawati Darma Astuti; Ahmad Ishaq; Sutrisno; Eva Argarini Pratama in [15] studied different clustering techniques.

Traditional segmentation methods often rely on manual categorization or simple rules, which limit their effectiveness in capturing the complexity of customer preferences and behaviors. In recent years, machine learning (ML) techniques have become prominent in customer segmentation due to their ability to handle large volumes of data and uncover intricate patterns not easily identifiable through traditional methods. Various studies have examined the application of different ML algorithms, including k-means clustering, hierarchical clustering, and Gaussian mixture models, in customer segmentation tasks. These algorithms provide flexibility and scalability, allowing businesses to identify meaningful customer segments and tailor their marketing strategies accordingly. Despite the advantages of ML-based segmentation, challenges such as data quality issues, algorithm selection, and interpretability constraints persist. In response to these challenges, researchers have proposed hybrid models that combine the strengths of ML techniques with traditional statistical methods to improve segmentation accuracy and interpretability. Hybrid models often integrate clustering algorithms with dimensionality reduction techniques or ensemble methods to enhance segmentation performance while maintaining transparency in the segmentation process. Furthermore, the literature showcases practical applications of ML-driven segmentation and hybrid models across various industries, including retail, e-commerce, banking, and telecommunications. Real-world case studies illustrate the effectiveness of these approaches in identifying distinct customer segments, predicting customer behavior, and optimizing marketing campaigns. In summary, the literature emphasizes the increasing importance of ML-driven

customer segmentation and the emergence of hybrid models as a promising approach to addressing the challenges associated with traditional segmentation methods. Future research directions may focus on developing more advanced hybrid models, addressing algorithmic biases, and exploring novel applications of ML techniques in customer segmentation to further enhance marketing effectiveness and customer satisfaction.

# III THEORETICAL BACKGROUND

## 3.1 PROBLEM IDENTIFICATION

- The existing model for the customer segmentation depicts that it is based on the K-means clustering algorithm which comes under centroid-based clustering. The suitable K value for the given dataset is selected appropriately which represents the predefined clusters. Raw and unlabelled data is taken as input which is further divided into clusters until the best clusters are found. Centroid based algorithm used in this model is efficient but sensitive to initial conditions and outliers.

## 3.2 PROBLEM SOLVING

- In the proposed system, the customer segmentation model includes not only centroid-based but also hierarchical clustering. • The three clustering algorithms K means, Minibatch K means and the hierarchical algorithm has been selected from the literature survey. • By deploying the three different algorithms, the clusters are formed and analysed respectively. • The most effective and efficient algorithm is determined by comparing and evaluating the precision rate among the three algorithms.

- To begin, we import the necessary packages to do our analysis and then the xlsx (Excel spreadsheet) data file.[12] If you want to follow up with the same data, you have to download it from UCI. For this example, I place the xlsx file in the folder (directory) where I present Jupiter's notebook.

- After importing the package and data, we will see that the data is not as helpful as that, so we need to clean and organize this data in a way that we can create more actionable insights.

- The K-means area unit is sensitive to the scale of the information used, such as clustering algorithms, so we would like to generalize the information.

## 3.3 SYSTEM ARCHITECTURE
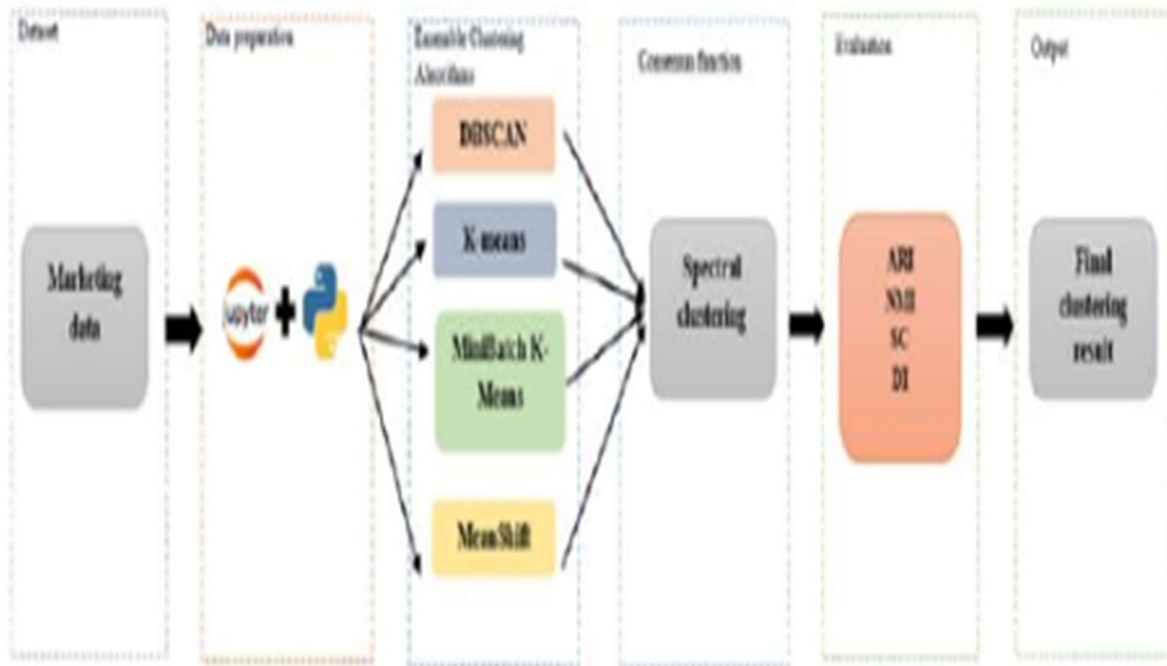


## IV SYSTEM IMPLEMETATION

### 4.1. Modules:

- Dataset Collection
- Train and test the model
- Deploy the models

### Modules DESCRIPTION:

- **Dataset Collection**- We had collected datasets from Kaggle notebooks. The dataset contains the symptoms and the corresponding disease. It contains 200 rows and 5 columns.
- **Train and test the model**- We had used three clustering algorithms named K-means clustering algorithm, Hierarchical clustering, and mini-batch K-means algorithm to train the dataset. After training, we had tested the model and found their clusters, silhouette score, and Davies Boulding score.

- **Deploy the models**- Deployed the model to get the clusters formed. The cluster shows the different segmentation of customers based on many attributes. By this, we will get the silhouette score and Davies Boulding scores of the model as the output.

## V CONCLUSION & FUTURE WORK

### 5.1 CONCLUSION

In this project, segments of customers are created using the k-means clustering model and analysed the dataset, in various ways. Visualization of the data set has been done for the better understanding about all the elements and its relation between the data. We used a clustering approach called K-means clustering, in particular. Kmeans clustering is one of the most popular clustering methods, and it's frequently the first thing practitioners try when they're working on a clustering problem. K- means are used to divide data points into discrete, nonoverlapping groupings. One of the most common uses of K-means clustering is client segmentation in order to gain a better understanding of them, which can then be used to boost the company's income.

### REFERENCE

[1] Aman Banduni, Prof Ilavedhan A, "Customer Segmentation using machine learning," School of Computing Science and Engineering, Galgotias University, Greater Noida, Uttar Pradesh, India. •

[2] Kamalpreet Bindra, Anuranjan Mishra, "A Detailed Study of Clustering Algorithms", CSE Department, Noida International University, In 2017 6th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), In September 2017, AIIT, Amity University Uttar Pradesh, Noida, India. [3] KAI PENG(Member, IEEE), VICTOR C. M. LEUNG, (Fellow, IEEE), AND QINGJIA HUANG, "Clustering Approach based on Mini batch Kmeans ", In 2018, College of Engineering, Huaqiao University, Quanzhou 362021, China.

[4] Fionn Murtagh and Pedro Contreras, "Methods of Hierarchical Clustering", In 2018, Science Foundation Ireland, Wilton Place, Dublin, Ireland Department of Computer Science, Royal Holloway, University of London

[5] D. P. Yash Kushwaha, Deepak Prajapati, "Customer Segmentation using KMeans Algorithm," 8th Semester Student of Beach in Computer Science and Engineering, Galgotias University, India.

[6] Manju Kaushik, Bhawana Mathur, "Comparative Study of K-Means and Hierarchical Clustering Techniques", In June 2014, JECRC University, Jaipur.

[7] Ali Feizollah, Nor Badrul Anuar, Rosli Salleh, and Fairuz Amalina, "Comparative Study of K-means and Mini Batch K-means Clustering Algorithms", Computer System and Technology Department, The University of Malaya, Kuala Lumpur, Malaysia, International Journal of Software and Hardware Research in Engineering.

[8] Asith Ishantha, "Mall Customer Segmentation using Clustering Algorithm", Future University Hakodate, Conference Paper, March 2021.

[9] Onur Dogan, Dokuz eylul University, Ejder Aycin, Kocaeli University, Zeki Atil Bulut, Dokuz Eylul University, "Customer Segmentation by using RFM model and clustering methods: A case Study in Retail Industry", In July 2018, International Journal of Contemporary Economics and Administrative Sciences.

[10] Juni Nurma Sari,Ridi Ferdiana,Lukito Nugroho,Paulus Insap Santosa,"Review on Customer Segmentation Technique",Department of Electrical Engineering and Information Technology, University of Gadjah Mada, Jogjakarta, Indonesia, Department of Informatics Technology, Polytechnic Caltex Riau, Pekanbaru, Indonesia