Customer segmentation using K-Mean Algorithm

MrsG.S.Geethamani[1] Mr C.S.Kumaravel[2] Mr M.Manojkumar[3]

[1]*Assistant Professor, Department of Computer Science, Hindusthan College of Arts & Science, Coimbatore*

[2]*II PG Student, Department of Computer Science, Hindusthan College of Arts & Science, Coimbatore*
[3]*II PG Student, Department of Computer Science, Hindusthan College of Arts & Science, Coimbatore*

Abstract: Customer segmentation became very popular method for dividing company's customers for retaining customers and making profit out of them, in the following study customers of different of organizations are classified on the basis of their behavioral characteristics such as spending and income; by taking behavioral aspects into consideration makes these methods an efficient one as compares to others. For this classification a machine algorithm named as k-means clustering algorithm is used and based on the behavioral characteristic's customers are classified. Formed clusters help the company to target individual customer and advertise the content to them through marketing campaign and social media sites which they are really interested in.

Keywords: Machine learning, Linear aggression, prediction.

I.INTRODUCTION

Any company in retail, no matter the industry, ends up collecting, creating, and manipulating 1 data over the course of their lifespan. These data are produced and recorded in a variety of contexts, most notably in the form of shipments, tickets, employee logs, and digital interactions. Each of these instances of data describes a small piece of how the company operates, for better or for worse. The more access to data that one has, the better the picture that the data can delineate. With a clear picture made from data, details previously unseen begin to emerge that spur new insights and innovations. The sheer size and complicated nature of data in the real world make the above task much easier said than done, though. The rise of performance metrics and interactive dashboard have ushered in a new era of looking at data. Many times, the data included in dashboards are at the superficial level: How much did store X make during December?, What are our top 5 products?, What is our monthly COGS (Cost of Goods Sold)?. While dashboards supply data that often have important significance in supply chain management and operations, they are limited in the sense that they omit data and insights that require higher level of datamining and analysis. Companies that utilize proper data science and data mining practices allow themselves to dig further into their own operating strategies, which in turn allows them to optimize their commercial practices.

II.LITRATURESURVEY

Customer segmentation is a crucial technique in marketing and business analytics to divide customers into distinct groups based on their characteristics, behaviors, and preferences. The K-means clustering algorithm is one of the most widely used methods for customer segmentation due to its simplicity, efficiency, and scalability. Below is a literature survey summarizing the application of K-means for customer segmentation in various research contexts.

1.K-Means Overview: K-means is an unsupervised machine learning algorithm that partitions data into KKK clusters based on the features of the data points. Each cluster is represented by the centroid, which minimizes the variance of points within that cluster.

2.Application in Customer Segmentation: K-means is commonly applied to segment customers based on attributes like demographic data (age, gender), behavioral data (purchase frequency, transaction amount), or psychographic data (preferences, motivations). The goal is to group customers who exhibit similar characteristics into clusters for targeted marketing and personalized services.

3.E-CommerceandRetailSector

Wuetal.(2021): Applied K-means clustering to segment online shoppers based on transaction data such as frequency, recency, and monetary value

(RFMmodel). This study found that using K-means clustering allowed for the identification of high-value customers and low-engagement users, which helped in tailoring marketing campaigns more effectively.

Kumar & Shah (2019): Focused on customer segmentation for ane-commerce platform, using K-means clustering with demographic and transaction data. The clusters identified varied significantly in terms of spending behavior, indicating the need for different promotional strategies for each customer segment.

Bose et al. (2020): Discussed how K-means was used to create customer personas in a retail context, where the algorithm was used to analyze purchase patterns and customer loyalty to develop targeted pricing strategies and product offerings.

III.K-MEANALGORITHM

K-Means is a popular unsupervised machine learning algorithm used for clustering. Clustering involves grouping a set of objects such that objects within a group (or cluster) are more similar to each other than to those in other groups. K-Means is widely used in a variety of applications, including customer segmentation, image compression, anomaly detection, and more.

Clusters: K-Means divides a dataset into KKK clusters, where each cluster contains data points that are similar to each other.

Centroids: Each cluster is represented by a centroid, which is the mean of all the data points in that cluster.

Distance Measure: K-Means generally uses Euclidean distance to measure the similarity between data points and centroids. However, other distance measures (like Manhattan distance) can also be used depending on the problem.

Customer Segmentation: In marketing, K-means is used to segment customers based on purchasing behavior, demographics, and other attributes.

Image Compression: K-means is used to reduce the number of colors in an image by clustering pixel colors.

Anomaly Detection: K-means can be used to detect anomalous points by identifying data points that do not fit well with any cluster.

Document Clustering: K-means is applied in text mining to cluster similar documents based on word frequency vectors.

IV. DATAANALYSISANDMODELLING

Customer segmentation is the process of dividing a customer base into distinct groups based on specific characteristics like demographics, behavior, and purchasing patterns. The goal is to enable targeted marketing, personalized experiences, and better business strategies. data analysis and modeling allow businesses to understand their customer base by identifying meaningful patterns. Using clustering algorithms like K-Means, you can divide customers into segments based on their behavior, demographics, and preferences. With Python, you can easily preprocess, model, and visualize customer data to uncover actionable insights, leading to better marketing strategies, customer satisfaction, and business growth.

V.RESULTANDANALYSIS

The results and discussion of customer segmentation using machine learning typically involve evaluating the effectiveness of the segmentation model and interpreting the customer segments generated. The evaluation metrics used depend on the specific machine learning algorithm used and the business problem being solved. The results are discussed to identify the strengths and weaknesses of the segmentation model and to suggest possible improvements. The discussion can also include recommendations for future research segmentation model

VI.CONCLUSION

Customer segmentation using machine learning is a powerful technique that can help businesses understand their customers better and tailor their marketing strategies accordingly. By segmenting customers based on their characteristics, businesses can improve customer engagement, increase sales revenue, and optimize their marketing efforts. The process involves collecting and analyzing customer data, selecting appropriate machine learning algorithms, and evaluating the effectiveness of the segmentation model

REFERRENCE

[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] Chinedu Pascal Ezenkwu, Simeon Ozuomba, Constance kalu Electrical/Electronics and Computer Engineering Department, University of Uyo, Akwa Ibom State, Nigeria "Application of K Means Algorithm for Efficient Customer Segmentation: A Strategy for Targeted Customer Services", IJARAI, Year: 2015.

[4] Potharaju, S. P., Sreedevi, M., Ande, V. K., & Tirandasu, R. K. (2019). Data mining approach for accelerating the classification accuracy of cardiotocography. Clinical Epidemiology and Global Health, 7(2), 160-164.

[5] Yogita Rani and Dr. Harish Rohil"A Study of Hierarchical Clustering Algorithm", IJICT, Year: 2013.