Team No 1

Abhinavtej Reddy - K Sumanth

Problem Statement: Customer Segmentation using Machine Learning

Algorithm Used: k-Means Clustering Algorithm

The k-Means clustering algorithm is a popular unsupervised machine learning technique used for partitioning data into clusters based on similarity. It aims to group data points into k clusters where each point belongs to the cluster with the nearest mean. In this report, we analyze the application of the k-Means algorithm on customer data from a mall.

Dataset Description

The dataset contains information about mall customers, including their age, gender, annual income, and spending score.

Choosing Features for Clustering

- The features chosen for clustering were "Annual Income" and "Spending Score.
- These features were selected as they are relevant for segmenting customers based on their purchasing behavior.

Determining the Optimal Number of Clusters

- The Elbow Method was utilized to determine the optimal number of clusters.
- A graph was plotted with the number of clusters ranging from 1 to 10.

- The inertia, which represents the sum of squared distances of samples to their closest cluster center, was used to assess cluster quality.
- Based on the Elbow Method, the optimal number of clusters was determined to be 5.

Training the k-Means Clustering Model

- The k-Means algorithm was trained with 5 clusters using the chosen features.
- The ‘k-means++’ initialization method was used to improve convergence speed.
- The model was fitted to the data, and cluster assignments were obtained for each data point.

Visualizing Clusters and Centroids

- Scatter plots were created to visualize the clusters formed by the k-Means algorithm.
- Each cluster was distinguished by a different color, and centroids were marked in cyan.
- The x-axis represented the "Annual Income," while the y-axis represented the "Spending Score."

Conclusion

- The k-Means algorithm successfully segmented mall customers into distinct groups based on their annual income and spending score.
- The visualization helped in understanding the distribution of customers across different clusters and their purchasing patterns.
- Further analysis and targeted marketing strategies can be developed based on the insights gained from these clusters.

Cluster 1: Average Annual Income & Average Spending Score

Target: This group represents customers with moderate income levels who spend moderately.

Suggestions: Offer a variety of mid-range products and services, focus on value for money, provide loyalty programs, and emphasize convenience and accessibility.

Cluster 2: High Annual Income & High Spending Score

Target: This group consists of affluent customers who have a high propensity to spend.

Suggestions: Create a luxurious shopping experience with high-end brands, personalized services, exclusive events, gourmet dining options, and VIP lounges to cater to their tastes and preferences.

Cluster 3: Low Annual Income & High Spending Score

Target: Despite having a low income, this group spends significantly.

Suggestions: Offer affordable luxury items, budget-friendly deals, installment plans, and loyalty programs. Focus on trendy and fashionable items at reasonable prices to attract this segment.

Cluster 4: Low Annual Income & Low Spending Score

Target: Customers in this group have limited income and spend conservatively.

Suggestions: Provide budget-friendly options, discounts, and promotions, emphasize the value and affordability of products, and offer practical and essential goods.

Cluster 5: Low Annual Income & High Spending Score

Target: This group represents customers who may be stretching their budgets to spend more.

Suggestions: Offer flexible payment options, discounts, and deals tailored to their spending patterns. Focus on providing value for money and emphasize the benefits of your products and services to justify their spending.