Afia Tasnim Ria

0112231058

K-Means Clustering for Customer Segmentation

Introduction

K-Means clustering is an unsupervised machine learning algorithm used to identify patterns and group similar data points. This technique is commonly applied in customer segmentation, where businesses categorize customers based on spending behavior and online activity. By segmenting customers, businesses can tailor their marketing strategies and services to different customer groups effectively.

1. Dataset Overview

The dataset consists of 10 customers, each described by two features:

- Monthly Spending (USD): The amount spent by the customer each month.

- Number of Website Visits: The frequency of website visits per month.

Customer Data

Customer ID	Monthly Spending (USD)	Number of Visits
C1	100	2
C2	250	5
C3	700	10
C4	50	1
C5	1800	3
C6	900	12
C7	120	7
C8	1600	2
C9	60	15

Customer ID	Monthly Spending (USD)	Number of Visits
C10	1100	8

2. Clustering Process

To segment customers, we apply K-Means clustering with K=3 (three clusters). The algorithm follows these steps:

1. Initialization: Select 3 initial cluster centroids randomly.

2. Assignment: Assign each customer to the nearest cluster center.

3. Update: Compute new cluster centroids by averaging the values of all customers assigned to each cluster.

4. Repeat: Steps 2 and 3 continue iteratively until the clusters stabilize and do not change significantly.

The Euclidean distance is used to measure similarity between data points and cluster centers.

3. Clustering Results

After running the algorithm, the customers are grouped into three distinct segments:

- Cluster 0: Low to moderate spenders with fewer visits.

- Cluster 1: Moderate to high spenders who visit more often.

- Cluster 2: Very high spenders who visit only a few times.

Visualization

A scatter plot is generated to visualize the clusters:

- Data points are colored based on cluster assignment.

- Red markers indicate the computed cluster centroids.

4. Predicting New Customers

Once the model is trained, we can predict the cluster for new customers based on their spending and website visit frequency.

Example Predictions

Spending (USD)	Visits	Predicted Cluster
300	4	Cluster 0
1500	1	Cluster 2

6. Conclusion

Using K-Means clustering, we successfully segmented customers based on their spending and website activity. This segmentation helps businesses understand customer behavior and improve engagement strategies. The model can be extended to larger datasets for deeper insights into customer preferences.