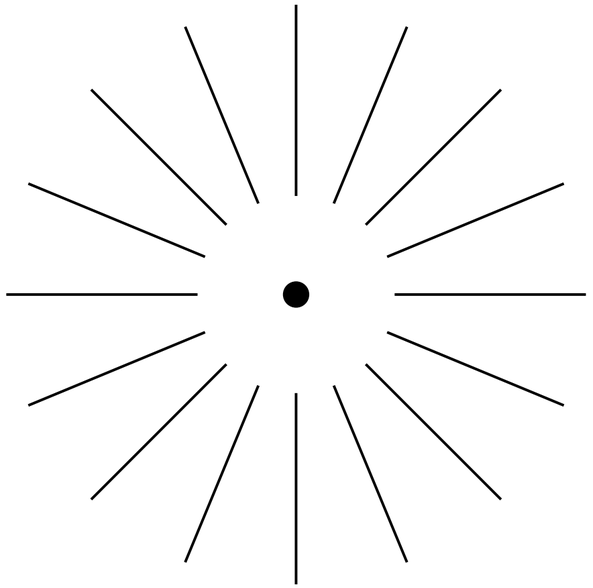
CUSTOMER SEGMENTATION CLASSIFICATION

SHIVAM MISHRA
SHRADDHESH BHALERAO
SIDDHARTH CHAUDHRY
VIDUSHEE GUPTA

CONTENT



INTRODUCTION:

Customer Segmentation is the process of division of customer base into several groups of individuals that share a similarity in different ways that are relevant to marketing such as gender, age, interests, and miscellaneous spending habits. Using clustering techniques, companies can identify the several segments of customers allowing them to target the potential user base. In this machine learning project, we will make use of k-mean Clustering which is the essential algorithm for clustering unlabeled dataset.

PROBLEM STATEMENT

Customer Segmentation is a crucial practice for businesses seeking to optimize their marketing strategies and tailor their products or services to different consumer groups. Unsupervised learning techniques, specifically clustering, offer a valuable approach to identify distinct customer segments based on their behaviors, preferences, and demographics.
The problem at hand is to explore the application of K-means Clustering, a fundamental algorithm for clustering unlabeled datasets, to enhance the marketing and sales strategies of a company

METHODOLOGY: PROCEDURES AND TOOLS

1. Project Objective:
   - Goal: Identify and categorize distinct customer segments based on their behavior and characteristics.
   - Method: Utilize K-means clustering, an unsupervised learning algorithm, to group customers with similar attributes.
2. Data Collection and Exploration:
   - Data Source: Gather relevant customer data, including features such as purchase history, frequency, and monetary value.
   - Exploration: Understand the dataset's structure, handle missing values, and perform exploratory data analysis (EDA).
3. Feature Selection:
   - Choose Relevant Features: Select features that capture meaningful aspects of customer behavior (e.g., frequency of purchases, total spending, etc.).
4. Data Preprocessing:
   - Scaling: Standardize numerical features to ensure equal influence during clustering.
   - Handling Categorical Data: If applicable, encode or transform categorical variables.
5. Determine Optimal Number of Clusters:
   - K-values Exploration: Use the elbow method by running K-means clustering for a range of k-values (number of clusters).
   - Sum of Squares (Inertia): Plot the sum of squared distances within clusters against different k-values. Identify the "elbow" point, where additional clusters provide diminishing returns.
6. K-means Clustering:
   - Model Training: Implement K-means clustering using the optimal k-value determined in the previous step.
   - Assigning Labels: Assign cluster labels to each customer based on their characteristics.
7. Interpretation of Clusters:
   - Cluster Profiles: Analyze the characteristics of each cluster to understand the distinct traits of customers within each group.
   - Insights: Derive insights into customer preferences, behaviors, and needs for each segment.
8. Validation:
   - Internal Validation: Assess the quality of clusters using metrics like silhouette score.
   - External Validation: If applicable, validate clusters against external benchmarks or expert knowledge.

K-MEANS ALGORITHM

While using the k-means clustering algorithm, the first step is to indicate the number of clusters
(k) that we wish to produce in the final output. The algorithm starts by selecting k objects from dataset randomly that will serve as the initial centers for our clusters. These selected objects are the cluster means, also known as centroids. Then, the remaining objects have an assignment of the closest centroid. This centroid is defined by the Euclidean Distance present between the object and the cluster mean. We refer to this step as “cluster assignment”. When the assignment is complete, the algorithm proceeds to calculate new mean value of each cluster present in the data. After the recalculation of the centers, the observations are checked if they are closer to a different cluster. Using the updated cluster mean, the objects undergo reassignment. This goes on repeatedly through several iterations until the cluster assignments stop altering. The clusters that are present in the current iterationare the same as the ones obtainedin the previous iteration.

APPLICATION

Optimal Cluster Determination:
The elbow point in the plot indicates the optimal number of clusters for customer segmentation.

Segmentation Insights:
Utilized the optimal number of clusters to perform k-means clustering on the entire dataset.
Extracted insights into distinct customer segments based on their characteristics (features 3 to 5).

Business Strategy Implementation:
Informed marketing strategies tailored to each customer segment.
Improved customer experience by addressing the unique needs and preferences of different segments.

Personalized Campaigns:
Designed targeted campaigns for each customer segment to enhance engagement and conversion rates.

Iterative Refinement:
Regularly updated the customer segmentation based on new data to adapt to changing trends and preferences.

REFERENCES

- Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.). Springer. - This book discusses machine learning and clustering methods in depth.
- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. Pattern Recognition Letters, 31(8), 651-666. - A seminal paper on the history and evolution of clustering techniques, including K-means
- R, S., & Murty, M. N. (1995). Automatic generation of K-means clustering algorithm. Information Sciences, 86(3-4), 181-195. - This paper explores automatic generation of K-means clustering algorithms.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability (Vol. 1, No. 14, pp. 281-297). - The original paper on K-means clustering by its creator, J. MacQueen.

THANK YOU VERY MUCH!