Customer Personality Analysis

Understanding Customer Segments Using Unsupervised Learning

What Problem Are We Solving?

-Businesses often struggle to understand their customers' behavior and preferences.

- Without proper segmentation, marketing efforts can be inefficient and costly.

Objective:

- Perform customer segmentation to identify distinct groups of customers based on their demographics, spending habits, and purchasing behavior.

- Tailor marketing strategies to each segment to improve engagement and sales.

Dataset Overview

Dataset: Customer Personality Analysis (Kaggle)

- Rows: 2,240 customers

- Columns: 29 features

- Key Features:

  - Demographics: Age, Income, Marital Status, Education, Number of Children

  - Spending: Amount spent on wine, meat, fish, sweets, etc.

  - Purchases: Store, web, catalog purchases

  - Promotions: Campaign responses

Goal: Use unsupervised learning to group customers into meaningful segments.

Exploratory Data Analysis

Key Insights:

- Income Distribution: Most customers earn between $20k and $80k.

- Spending Habits: Wine and meat are the most purchased categories.

- Correlation: Income is positively correlated with spending on luxury items like wine and gold products.

Machine Learning Approach

Unsupervised Learning Models Used:

1. Principal Component Analysis (PCA):

  - Reduced dimensionality to capture 90% of variance with fewer components.

  - Visualized clusters in 2D using the first two principal components.

2. Clustering Algorithms:

  - K-Means: Optimized for compact, spherical clusters.

  - Gaussian Mixture Model (GMM)**: Probabilistic clustering for overlapping clusters.

  - Agglomerative Clustering: Hierarchical clustering to capture relationships.

Evaluation Metric:

- Silhouette Score: Measures cluster cohesion and separation.

Results

Cluster Analysis:

- Number of Clusters: 4 (based on the Elbow Method).

- Silhouette Scores:

  - K-Means: 0.2489 (Best)

  - GMM: 0.0840

  - Agglomerative Clustering: 0.2226

Interpretation:

- K-Means achieved the best cluster separation.

- GMM struggled due to overlapping clusters.

- Agglomerative Clustering performed reasonably well but was less distinct than K-Means.

Recommendations

Marketing Strategies:

1. Group 0 (Moderate Earners):

- Highlight variety and convenience.

- Promote balanced product bundles.

2. Group 1 (Lower Income):

- Offer value deals and discounts.

- Focus on affordable products.

3. Group 2 (Premium Buyers):

- Promote premium products and loyalty programs.

- Personalized offers for high-value customers.

4. Group 3 (Low Engagement):

- Re-engagement campaigns.

- Focus on budget-friendly options.

Conclusions

Summary:

- Successfully segmented customers into 4 distinct groups using unsupervised learning.

- K-Means provided the best clustering results.

- Insights from clusters can guide targeted marketing strategies.

Next Steps:

- Validate clusters with additional data.

- Explore supervised learning for predicting customer behavior.

- Implement marketing strategies and measure ROI.

Thank You