Project Proposal: Customer Segmentation Using Machine Learning

Objective

The objective of the project is to develop a machine learning model that segments customers into four predefined groups, namely A, B, C, and D, based on behavioral patterns from existing market data. Further, this will enable focused marketing and communication strategies for 2627 new potential customers in a new market with the learning from the existing market.

Problem Statement

A car company wants to expand into new markets with its range of products, P1, P2, P3, P4, and P5. From their current market, they have historical sales data that proves the success of segment-specific marketing strategies based on the division of customers into four groups. For this to work in the same way, the company needs a robust machine learning solution to predict the segment of new customers by finding out how similar they are to the already existing customers.

Scope of the Project

The project involves:

1. The analysis of the dataset for customer behaviors of the existing market.
2. The construction and validation of the machine learning model to segment customers into predefined segments.
3. Provide insightful knowledge to the sales team for personalized outreach into the new market.

Dataset Description

The dataset includes information about customer behaviors in the existing market, segmented into four groups: A, B, C, and D. The company has also provided data for 2627 new potential customers to be classified.

The dataset was sourced from Kaggle: [Customer Segmentation Dataset](https://www.kaggle.com/datasets/vetrirah/customer).

Deliverables

1. Exploratory Data Analysis (EDA):

* Comprehensive exploration of the dataset, including handling missing values, visualizations, and feature engineering.

1. Machine Learning Model:

* Application of clustering methodologies like K-Means for customer segmentation.
* Model validation and performance evaluation, including the use of silhouette scores.

1. Final Report:

* Customer outreach: insights and recommendations.
* Overall Documentation of Workflow: EDA, modelling and results.

1. Code Implementation:

* A reproducible Jupyter Notebook with all the code for data cleaning, EDA, modeling, and predictions.

Methodology

1. Data Preprocessing:

* Handling missing values and outliers.
* Normalizing features to make a better clustering performance.

1. Exploratory Data Analysis:

* Identifying patterns and trends within the data.
* Visualize feature distributions and segment characteristics.

1. Model Development:

* Train K-Means clustering algorithm to detect the segments of customers.
* Optimize number of clusters and validate using Silhouette Analysis.

1. Prediction and Insights:

* Predict the segments for these 2627 new customers
* Providing actionable insights for segment-specific outreach.

Tools and Technologies

• Python Libraries: Pandas, NumPy, Matplotlib, Seaborn, scikit-learn.

• Environment: Jupyter Notebook.

• Clustering Algorithm: K-Means.

Expected Outcome

The project aims to deliver:

1. A machine learning solution will be developed that can appropriately predict customer segments of the new market.
2. Clear, data-driven insights to support the company’s marketing strategy.
3. Documentation and codebase for reproducibility and future scalability.

Conclusion

This project will bridge the gap in existing market data and new market entry, thus helping the company to retain its lead in segment-specific customer engagement. The company is going to dive deeper into understanding its customer base and growing in the market in a data-driven way with machine learning.