Customer Personality Analysis

Machine learning project

Document Type : Detailed Project Report (DPR)

Author : Nisha Doshi

Date : 06-02-2025

1. Project Overview

Customer segmentation is crucial for businesses to understand their audience, enhance customer experiences, and optimize marketing strategies. This project focuses on segmenting customers based on their purchasing behaviors using clustering techniques. The insights help in targeted marketing strategies to improve business efficiency.

2. Objectives

- To analyze customer purchasing behaviors and spending patterns.
- To segment customers into distinct groups using clustering algorithms.
- To provide actionable insights for targeted marketing strategies.
- To develop a web-based application using Flask for interactive analysis and visualization.

3. Data Description

- **Source**: https://raw.githubusercontent.com/amankharwal/Website-data/master/marketing_campaign.csv
- **Key Features**:
  - **Demographic Information**: Age, Marital Status, Education, Income
  - **Behavioral Data**: Number of purchases, Total amount spent, Complaints
  - **Marketing Response**: Accepted campaigns
- **Preprocessing Steps**:
  - Handling missing values (Income column replaced with median values)
  - Removing duplicates
  - Encoding categorical variables (Education & Marital Status)
  - Feature engineering:
    - Creating "Age" from "Year_Birth"
    - Merging education categories (Basic & 2nd Cycle → School)
    - Summarizing kids' data (Teenhome + Kidhome)
    - Aggregating total spending & purchase numbers
  - Outlier treatment using IQR method (Age > 90 removed, Income outliers handled)

4. Methodology

4.1 Exploratory Data Analysis (EDA)

EDA is performed using Matplotlib, Seaborn, and Plotly to visualize:

- Spending behavior trends
- Income distribution
- Relationship between income and total spending
- Purchase trends
- Effectiveness of marketing campaigns

4.2 Clustering Techniques

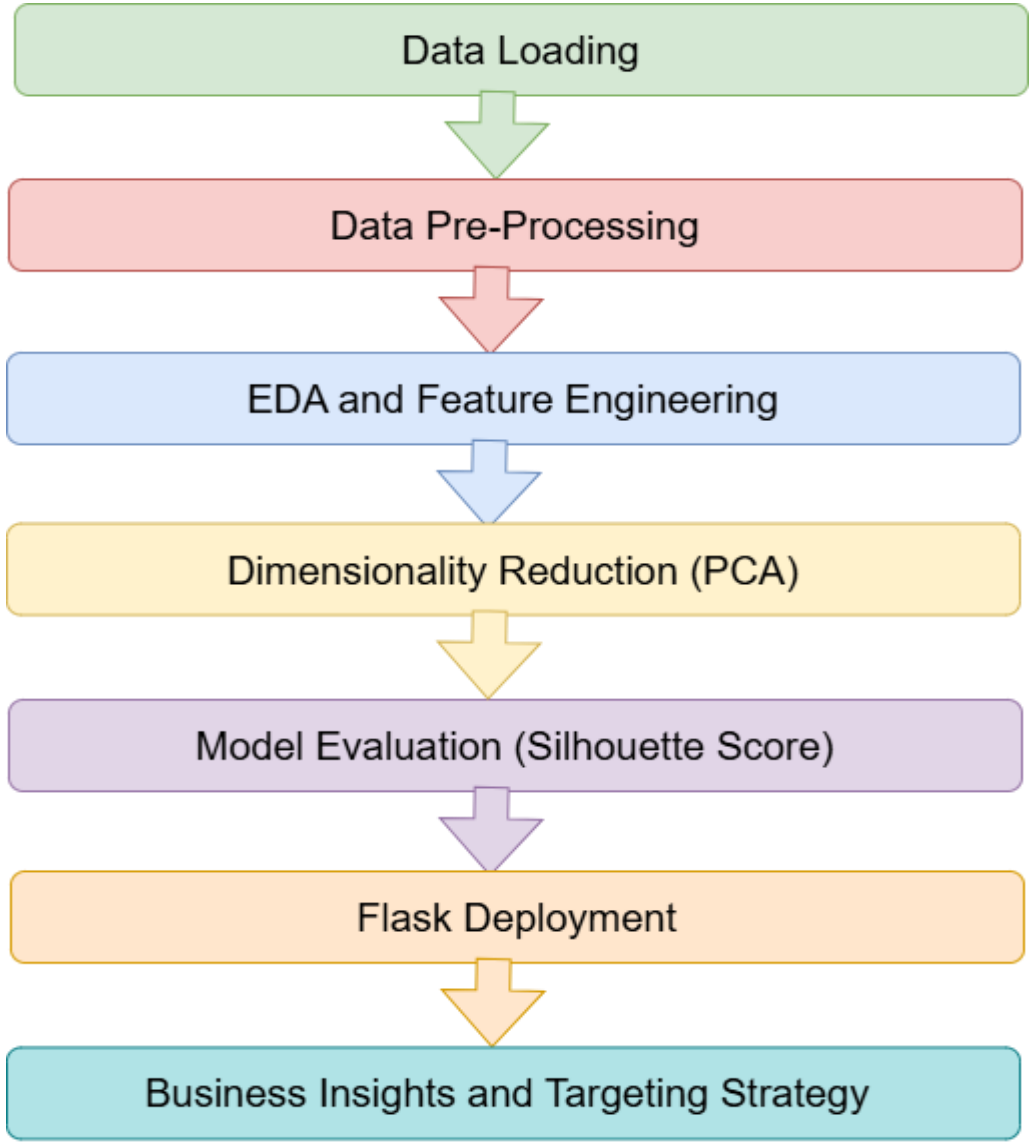
- **Dimensionality Reduction**: PCA to reduce feature space
- **Clustering Algorithms**:
  - K-Means Clustering (Optimal number of clusters determined using Elbow Method & Silhouette Score)
  - Hierarchical Clustering for additional insights
- **Evaluation Metrics**:
  - Silhouette Score to assess cluster quality

5. System Architecture

The system follows a modular design:

1. **Data Ingestion Layer**:
    1. Data is uploaded via a CSV file.
    2. Preprocessing pipeline executes missing value handling, encoding, and feature engineering.
2. **Processing Layer**:
    1. EDA is performed with visualization libraries.
    2. Clustering models are trained using K-Means and Hierarchical Clustering.
3. **Application Layer**:
    1. Web application developed using Flask for real-time interaction.
    2. Users can visualize data, apply clustering, and analyze insights.
4. **Deployment & API**:
    1. Flask API provides real-time predictions for customer segmentation.

Architecture



6. Web Application Features

6.1 Dashboard

- Displays an overview of customer segmentation insights.
- Shows key statistics such as total customers, average spending, and segmentation summary.

6.2 Data Upload & Processing

- Allows users to upload CSV files for analysis.
- Provides preprocessing summaries and automated data cleaning.

6.3 Data Visualization

- Interactive charts for income distribution, purchase trends, spending behavior.
- Cluster distribution analysis using PCA projections.

6.4 Clustering & Insights

- Optimal cluster determination using Elbow Method & Silhouette Score.
- Displays cluster characteristics and recommendations for business strategies.

6.5 Customer Prediction Tool

- Users can enter customer details to predict segmentation.
- Provides targeted marketing recommendations based on the assigned cluster.

7. Business Insights & Recommendations

7.1 Key Findings

- Customers spend more on wine and meat products.
- Physical store purchases are preferred over online purchases.
- Marketing campaigns have low acceptance rates.
- Senior and educated customers complain the most.

7.2 Cluster-Based Targeting Strategy

- **Cluster 1**: High-income, high-spending customers → Targeted with premium products & loyalty programs.
- **Cluster 2**: Budget-conscious customers → Targeted with discounts & affordable product bundles.

8. Conclusion & Future Work

- **Conclusion**: The project successfully segments customers and provides actionable insights for better marketing strategies.
- **Future Enhancements**:
  - Incorporate deep learning models for customer behavior prediction.
  - Integrate real-time customer segmentation for personalized recommendations.
  - Expand dataset to include geographical and behavioral data for better insights.