Customer Personality Analysis

MACHINE LEARNING PROJECT

DOCUMENT TYPE : LOW LEVEL DESIGN (LLD)

AUTHOR : NISHA DOSHI

DATE : 06-02-2025

1. Introduction:

THIS DOCUMENT PROVIDES A DETAILED TECHNICAL IMPLEMENTATION OF THE **CUSTOMER PERSONALITY ANALYSIS** PROJECT. THE PROJECT AIMS TO SEGMENT CUSTOMERS BASED ON THEIR PURCHASING BEHAVIOR AND DEMOGRAPHICS USING CLUSTERING TECHNIQUES.

PROBLEM STATEMENT

CUSTOMER PERSONALITY ANALYSIS IS A DETAILED STUDY OF A COMPANY'S IDEAL CUSTOMERS. IT HELPS BUSINESSES BETTER UNDERSTAND THEIR CUSTOMER BASE, ENABLING THEM TO TAILOR PRODUCTS AND MARKETING STRATEGIES ACCORDING TO CUSTOMER NEEDS, BEHAVIORS, AND CONCERNS.

INSTEAD OF MARKETING A NEW PRODUCT TO EVERY CUSTOMER, BUSINESSES CAN SEGMENT THEIR CUSTOMERS AND FOCUS MARKETING EFFORTS ON THOSE MOST LIKELY TO MAKE A PURCHASE. THE PRIMARY OBJECTIVES OF THIS PROJECT ARE:

1. UNDERSTANDING CUSTOMER ATTITUDES: ANALYZING WHAT CUSTOMERS SAY ABOUT THE PRODUCT TO UNDERSTAND THEIR SENTIMENTS AND OPINIONS.

2. ANALYZING CUSTOMER ACTIONS: IDENTIFYING CUSTOMER BEHAVIORS TO DETERMINE HOW THEY ENGAGE WITH THE PRODUCT.

2. Data Understanding & Preprocessing

2.1 DATASET SOURCE

- THE DATASET IS OBTAINED FROM KAGGLE.
- IT CONTAINS CUSTOMER DEMOGRAPHIC DETAILS, PURCHASE HISTORY, AND CAMPAIGN RESPONSES.

2.2 DATA CLEANING

- NO DUPLICATE VALUES IN THE DATASET.
- **HANDLING MISSING VALUES:** THE 'INCOME' COLUMN HAD MISSING VALUES, WHICH WERE REPLACED WITH THE MEDIAN.
- **DROPPING UNNECESSARY COLUMNS:** REMOVED COLUMNS: 'ID', 'Z_COSTCONTACT' AND 'Z_REVENUE' AS THEY DO NOT CONTRIBUTE TO ANALYSIS

2.3 FEATURE ENGINEERING

- CREATED AGE COLUMN FROM YEAR_BIRTH.

- MODIFIED EDUCATION CATEGORIES: REPLACED 'BASIC' AND '2N CYCLE' WITH 'SCHOOL' IN EDUCATION COLUMN.

- CREATED KIDS COLUMN: SUMMED UP TEENHOME AND KIDHOME.

- CREATED TOTAL_AMOUNT_SPENT COLUMN: SUMMED SPENDING AMOUNTS ACROSS ALL PRODUCT CATEGORIES (MNTWINES, MNTFRUITS, ETC.).

- CREATED NUMBER_OF_PURCHASES COLUMN: SUMMED NUMBER OF PURCHASES ACROSS DIFFERENT CHANNELS (NUMDEALSPURCHASES, NUMWEBPURCHASES, ETC.).

- CREATED TOTAL_ACCEPTED_CAMPAIGN COLUMN: SUMMED UP ALL CAMPAIGN ACCEPTANCE COLUMNS (ACCEPTEDCMP1 TO ACCEPTEDCMP5).

2.4 HANDLING OUTLIERS

- REMOVED CUSTOMERS AGED ABOVE 90.

- REMOVED OUTLIERS FROM INCOME COLUMN USING IQR METHOD.

2.5 ENCODING CATEGORICAL VARIABLES

- USED LABEL ENCODING FOR MARITAL_STATUS AND EDUCATION COLUMNS.

3. Exploratory Data Analysis (EDA)

- VISUALIZATION TOOLS: PLOTLY, SEABORN, MATPLOTLIB

- KEY FINDINGS:

  CUSTOMERS SPEND MORE ON WINE AND MEAT PRODUCTS.

  MORE PURCHASES ARE MADE FROM PHYSICAL STORES.

  CUSTOMERS DO NOT ENGAGE MUCH WITH MARKETING CAMPAIGNS.

4. Feature Selection

- SELECTED FEATURES: AGE, EDUCATION, MARITAL_STATUS, KIDS, INCOME, TOTAL_AMOUNT_SPENT, TOTAL_ACCEPTED_CAMPAIGN, NUMBER_OF_PURCHASES, COMPLAIN

- EXCLUDED FEATURES: RECENCY (NO SIGNIFICANT TREND), DAYS_FOR_CUSTOMER (NO MEANINGFUL PEAKS OR DROPS)

5. Clustering Methodology

- **DIMENSIONALITY REDUCTION:** USED PCA (PRINCIPAL COMPONENT ANALYSIS) TO REDUCE DIMENSIONALITY.

- **CLUSTERING ALGORITHMS:** APPLIED K-MEANS AND HIERARCHICAL CLUSTERING.

- **EVALUATION METRIC:** USED SILHOUETTE SCORE FROM SKLEARN.METRICS TO EVALUATE CLUSTERING PERFORMANCE.

6. Model Implementation

6.1 CODE BREAKDOWN

- **DATA LOADING**

- **PREPROCESSING PIPELINE**

- **CLUSTERING MODEL TRAINING**

- **HYPERPARAMETER TUNING**

- **MODEL EVALUATION**

7. Deployment Strategy

- **USED FLASK FOR DEPLOYMENT** TO CREATE A SIMPLE WEB INTERFACE.
- **API ENDPOINTS:**

   /PREDICT: ACCEPTS CUSTOMER INPUT AND RETURNS THEIR CLUSTER GROUP.

- **INTEGRATION:** CAN INTEGRATE WITH A POWER BI/TABLEAU DASHBOARD FOR VISUALIZATION.

8. Results & Insights

CUSTOMER SEGMENTS IDENTIFIED

1. **CLUSTER 1: HIGH-INCOME CUSTOMERS WITH HIGH SPENDING.**
2. **CLUSTER 2: LOW-INCOME CUSTOMERS WITH LOWER SPENDING.**
3. **BEST MODEL: K-MEANS CLUSTERING.**

KEY FINDINGS:

- HIGH-INCOME CUSTOMERS TEND TO SPEND MORE ON PREMIUM PRODUCTS.

- BUDGET-CONSCIOUS CUSTOMERS PREFER DISCOUNTS AND PROMOTIONS.

- MARKETING CAMPAIGNS ARE NOT EFFECTIVE, AS THEY RECEIVE LOW ENGAGEMENT.

9. Targeting Strategy

- CLUSTER 1: TARGET WITH PREMIUM PRODUCTS & LOYALTY PROGRAMS.

- CLUSTER 2: TARGET WITH DISCOUNTS & BUDGET-FRIENDLY OFFERS.

10. Conclusion

- CUSTOMERS PREFER PHYSICAL STORES OVER ONLINE SHOPPING.

- WINE AND MEAT PRODUCTS ARE THE MOST POPULAR.

- MARKETING CAMPAIGNS NEED BETTER ENGAGEMENT STRATEGIES.