Customer Behavior Analytics and Predictive Segmentation for Targeted Marketing

Ahmed Bilal Nazim
2022064
Department of Data Science
GIKI
Email: u2022064@giki.edu.pk

Ahmad Ali Khan
2022054
Department of Data Science
GIKI
Email: u2022054@giki.edu.pk

Abstract—This project applies a complete data mining pipeline to the Portuguese Marketing Campaign dataset to uncover customer behavior trends and enhance marketing strategy. Using end-to-end data preprocessing, exploratory data analysis, association rule mining, classification algorithms, clustering, and actionable insights, the study identifies key features influencing campaign success. This detailed analysis leads to predictive segmentation and strategic recommendations tailored for customer engagement and conversion optimization.

I. INTRODUCTION

In today's data-driven landscape, marketing departments strive to gain a deep understanding of customer behavior in order to optimize targeting and engagement. With abundant customer data being generated, it becomes essential to apply data mining techniques to extract actionable patterns and predictive insights. This project focuses on applying comprehensive data analytics on a real-world marketing campaign dataset to build models that can identify valuable customer segments and predict campaign responses. The project leverages data transformation, machine learning, and pattern discovery to support smarter decision-making in marketing strategy.

II. DATASET OVERVIEW

The dataset used in this study is a real-world dataset obtained from a Portuguese retail company, containing 2,240 records and 29 features. It includes customer demographic information (such as income, education, marital status), past product purchases (in wine, meat, fruits, etc.), and response to five different marketing campaigns. The target variable is 'Response', indicating whether a customer accepted the final campaign offer. The dataset includes both categorical and numerical variables, and required transformation before analysis.

III. DATA PREPROCESSING

A. Missing Values Handling

The dataset had 24 missing values in the 'Income' column. These were removed to maintain the integrity of the model training process.

B. Feature Transformation

The 'Dt_Customer' column was converted to datetime format, and a new feature 'Customer_Since_Days' was engineered to quantify customer loyalty in terms of duration.

C. Encoding and Normalization

Categorical variables such as 'Education' and 'Marital_Status' were transformed using one-hot encoding. All numeric features used in clustering and classification were standardized using Z-score normalization.

IV. EXPLORATORY DATA ANALYSIS (EDA)

EDA was conducted to explore trends and outliers. Histograms and boxplots of spending habits revealed that most customers spend significantly more on wine and meat than on other products. Correlation heatmaps showed positive relationships between income and purchasing behavior. Response rates were imbalanced, with fewer customers accepting the campaign offer. This required special attention during modeling.

V. ASSOCIATION RULE MINING

The Apriori algorithm was applied to binary-transformed purchase data. Key findings included:

- Rules like *[Gold, Meat, Fish] → [Fruits, Sweets]* had high confidence (¿0.80) and lift (¿1.10).
- These co-occurrence patterns are valuable for product bundling and cross-promotion.

Rules were visualized using scatter plots and 3D plots showing relationships between support, confidence, and lift.

VI. CLASSIFICATION MODELING

Three machine learning models were trained to predict the 'Response' variable:

- **Decision Tree:** Accuracy of 81.7%, suitable for interpretability but suffered from lower precision.
- **Naive Bayes:** Good recall (0.51), making it suitable when minimizing false negatives.
- **K-Nearest Neighbors (KNN):** Delivered highest accuracy (86.4%) and precision (0.53), making it the preferred model for deployment.

A performance comparison table was included, with metrics such as F1-score, recall, and precision.

VII. Customer Segmentation via Clustering

A. Clustering Technique

K-means clustering was performed on normalized behavioral features. PCA was used for dimensionality reduction to visualize cluster separability.

B. Cluster Profiles

- **Cluster 0:** Low-income, low-spend customers, low recency
- **Cluster 1:** High-income, loyal customers with high spending on wines and meats
- **Cluster 2:** Moderate-income, moderately engaged group

These segments were visualized in both 2D and 3D plots.

VIII. Actionable Insights and Recommendations

A. Insight Summary

- **Segment-Based Targeting:** Cluster 1 should be targeted with premium product campaigns.
- **Re-engagement Strategy:** Cluster 0 can be re-targeted using discount offers and personalized messages.
- **Cross-Selling Potential:** Association rules show high lift in co-purchases among wine, meat, and gold products. These combinations can be bundled.
- **Predictive Scoring:** Use KNN model to score new leads and prioritize customers likely to accept campaign offers.

B. Business Recommendations

- Invest in a recommendation engine based on the mined rules.
- Segment email campaigns by cluster profile.
- Monitor recency and campaign fatigue to re-engage disengaged customers.
- Automate customer scoring with the classification model.

IX. Challenges and Reflections

This project presented several challenges:

- **Data Parsing:** The dataset initially loaded as a single column and required correct delimiter handling.
- **Feature Engineering:** Converting dates and encoding categories required precision.
- **Model Selection:** Precision-recall tradeoffs were managed through metric comparison.
- **Interpretation:** Translating statistical output into actionable business decisions required critical thinking.

We learned how to move from raw data to valuable insights, integrating technical and business perspectives.

X. Conclusion

Data mining offers immense value to marketers by uncovering patterns that drive engagement and conversion. This project demonstrated how association rule mining, classification, and clustering can be combined to develop an effective, data-driven marketing strategy. Our findings highlight the importance of segmentation, targeted messaging, and predictive modeling in achieving successful campaign outcomes.

References

[1] Agrawal R., Srikant R., "Fast Algorithms for Mining Association Rules," Proc. VLDB, 1994.
[2] Scikit-learn: Machine Learning in Python, https://scikit-learn.org/
[3] Mlxtend: Python library for association rule mining, https://rasbt.github.io/mlxtend/