

Personalized Offer Response Prediction Starbucks

Project Proposal

1. Domain Background

In today's competitive retail market, personalization plays a crucial role in improving customer engagement, increasing sales, and optimizing marketing costs. Starbucks, like many other businesses, delivers targeted offers to its customers via its mobile app, including discounts, buy-one-get-one (BOGO), and informational promotions. Understanding which customer segments are most responsive to different types of offers is critical for increasing the return on investment of marketing campaigns.

This project is situated within the field of personalized marketing and customer behavior modeling. Predicting customer response to offers allows businesses to allocate marketing resources more efficiently. This type of problem sits at the intersection of data science, marketing analytics, and machine learning.

Machine learning enables us to analyze large, complex customer datasets and uncover patterns that would be difficult to detect. Accurate predictive models can improve business outcomes by increasing offer acceptance rates while reducing unnecessary costs associated with ineffective promotions.

2. Problem Statement

The objective of this project is to analyze customer response patterns to different types of marketing offers (BOGO, discount, informational) and identify which demographic segments are most responsive to each offer type.

Specifically, given the following:

- User's demographics (age, gender, income, membership start date)
- The details of a marketing offer (type, reward, difficulty, duration, channels, id)
- User's historical behavior (person, event, value, time)

We aim to build a predictive framework that estimates the likelihood of offer completion for each type of offer. This problem can be framed as a multi-label binary classification task, where:

- For each offer type, we predict whether the customer is likely to complete the offer when received.
- Results may then be aggregated to generate customer segmentations and targeting strategies.

3. Datasets and Inputs

The data used in this project is provided by Udacity's Starbucks Capstone Challenge dataset, which simulates customer behavior on the Starbucks mobile rewards app. The dataset includes three main JSON files:

profile.json

program users (17000 users x 5 fields)

- gender: (categorical) M, F, O, or null
- age: (numeric) missing value encoded as 118
- id: (string/hash)
- became_member_on: (date) format YYYYMMDD
- income: (numeric) Rewards

portfolio.json

Offers sent during 30-day test period (10 offers x 6 fields)

- reward:(numeric) money awarded for the amount spent
- channels:(list) web, email, mobile, social
- difficulty:(numeric) money required to be spent to receive reward
- duration:(numeric) time for offer to be open, in days
- offer type: (string) bogo, discount, informational
- id: (string/hash)

transcript.json

Event log (306648 events x 4 fields)

- person: (string/hash)
- event: (string) offer received, offer viewed, transaction, offer completed
- value: (dictionary) different values depending on event type

- offer id: (string/hash) not associated with any "transaction"
- amount:(numeric) money spent in "transaction"
- reward: (numeric) money gained from "offer completed"
- time: (numeric) hours after start of test

How the datasets will be used in the project:

- Join and merge these files appropriately.
- Clean missing or invalid entries.
- Derive meaningful features such as customer tenure, average transaction value, offer view status, and offer validity windows.

4. Solution Statement

The proposed solution involves building a supervised machine learning work that:

- Performs data cleaning and feature engineering to prepare the dataset.
- Encodes relevant categorical variables and temporal features.
- Builds a binary classification model to predict offer completion.
- Optionally explore multi-label models if time permits.

Business insights of this project are:

- Segment customer groups based on predicted responsiveness.
- Derive practical recommendations for offer targeting based on model results.

5. Benchmark Model

The benchmark model will be a simple **Logistic Regression** classifier using basic demographic features. This model will serve as a baseline to evaluate whether advanced machine learning models provide significant improvement.

6. Evaluation Metrics

Model performance will be evaluated using appropriate classification metrics:

- **Precision & Recall:** Balance between false positives and false negatives.
- **F1-score:** Harmonic mean of precision and recall.

Since the task is formulated as a binary classification problem with potential class imbalance, model performance will be evaluated using **F1-Score**. It allows us to focus on identifying responsive customers accurately while avoiding unnecessary marketing costs due to incorrect predictions.

Optionally, ROC-AUC may be used during model development to assess overall classifier discrimination ability, but F1-score will be used mainly for final model selection and reporting.

7. Project Design

Step 1: Data Preparation

- Load and merge the three data sources.
- Handle missing values.
- Process timestamps into meaningful features.
- Transform event logs into offer lifecycle stages.

Step 2: Exploratory Data Analysis (EDA)

- Visualize offer response rates across demographic groups.
- Analyze purchasing behavior with and without offers.
- Explore correlations between customer attributes and offer completions.

Step 3: Feature Engineering

- Generate features for:
 - Customer transaction history
 - Offer characteristics
 - Interaction behavior
 - Demographic variables

Step 4: Model Development

- Train initial logistic regression as benchmark.
- Train advanced models (Random Forest, XGBoost, LightGBM).
- Apply hyperparameter tuning

Step 5: Model Evaluation & Interpretation

- Use cross-validation for robust performance estimation.
- Analyze evaluation matrices and feature importances.
- Compare results across different offer types.

Step 6: Business Recommendations

- Translate model outputs into actionable marketing strategies.
- Present segmentation insights: which customer groups respond best to each offer.

Step 7: Final Reporting

- Prepare full documentation including:
 - Data processing steps
 - Modeling approach
 - Evaluation results
 - Visualizations
 - Business recommendations

References

- Starbucks Capstone Challenge Dataset (Udacity)
- Scikit-learn, XGBoost and LightGBM Documentation
- Kaggle marketing response prediction resources