

Starbucks Capstone Challenge Proposal

1. Domain Background

The project domain revolves around customer behavior analytics in the context of the Starbucks rewards mobile app. The goal is to study how customers react to promotional offers based on their unique demographics and historical interactions with the offers.

By understanding the customer responses, Starbucks can optimize its marketing strategies, enhance customer engagement, and boost revenue. The main objective of this Capstone Challenge is to dig deep into customer behavior to identify the most effective promotional offers for different customer segments.

2. Problem Statement

The primary goal is to identify the demographic profiles that respond positively to specific promotional offers by Starbucks. As Starbucks sends out a variety of offers to its rewards mobile app users, the response rate may differ across customer segments.

By gaining insights into customer demographics, and preferences, Starbucks can implement more targeted marketing efforts that enhance customer engagement and increase sales.

3. Datasets and Inputs

Three JSON datasets are provided as part of the challenge, containing a wealth of information about rewards program users, offers, and user interaction events:

- Profile dataset: Demographic information for 17,000 rewards program users. It includes age (average age of 62.53 years), became_member_on (2013 to 2018), income (average income of \$65,404, with a standard deviation of \$21,598).
- Portfolio dataset: Details for 10 offers sent during a 30-day test period. This dataset includes rewards ranging between 0 and 10, consisting of three types of offers: BOGO, discount, and informational.
- Transcript dataset: An event log containing 306,534 user interactions. It includes timestamp data (0 to 714 hours), transaction amounts (average transaction amount of \$12.78, with a standard deviation of \$30.25), and rewards (ranging between 2 and 10, with an average value of 4.9 and a standard deviation of 2.89).

4. Solution Statement

To address the problem, I propose developing a machine learning model to predict the likelihood of a user responding positively to a given promotional offer, considering their demographic information and past interaction history. The model will make use of customer data to estimate which specific promotional offer will be most effective for a certain customer segment. Amazon SageMaker's AutoGluon can be utilized for automatic model selection and hyperparameter tuning, selecting the best performing model from a variety of candidates, such as logistic regression, decision trees, and ensemble models.

5. Benchmark Model

A simple heuristic benchmark can be established by analyzing the historical response rate data for different demographic groups and offer types. This rule-based approach provides a basic insight into the effectiveness of various offers for each customer segment but lacks the refinement and accuracy of a machine learning model that considers a broader range of features and individual user interaction history.

6. Evaluation Metrics

To evaluate the model's performance in predicting user responses, the following evaluation metrics will be employed:

- **Accuracy:** Represents the proportion of correct predictions, considering both true positives and true negatives. In the context of our problem, it measures the overall success rate of predicting user responses to promotional offers.
- **Precision:** Measures the ratio of true positive predictions among all the positive predictions made. It reflects the model's ability to correctly predict which customers are likely to respond positively to a given offer.
- **Recall:** Calculates the proportion of actual positive instances that the model accurately predicted as positive. In our case, it measures the model's ability to capture all the customers that are truly responsive to an offer.
- **F1-Score:** The harmonic mean of precision and recall, delivering a balanced view of the model's performance.

For our specific use case, a higher F1-Score is desired to balance the trade-offs between precision and recall, ensuring the model's effectiveness in targeting the right customers with different promotional offers.

7. Project Design

The project will consist of the following key steps:

1. **Data Wrangling:** Clean and preprocess the input datasets, manage missing values, and format the data for subsequent analysis and modeling.
2. **Exploratory Data Analysis:** Perform both visual and statistical analysis to discover trends, correlations, and candidate features for the machine learning model.
3. **Feature Engineering:** Create or modify features to enhance the model's predictive performance.
4. **Data Splitting:** Split the prepared dataset into training and testing subsets.
5. **Model Selection:** Use SageMaker AutoGluon for automated model selection, identifying the best-performing model by comparing multiple algorithms and evaluations on the validation set.
6. **Model Training and Hyperparameter Tuning:** Train the selected model with the training set and fine-tune its hyperparameters using SageMaker's built-in tuning capabilities.
7. **Model Evaluation:** Evaluating the performance of the trained model on the test set using the chosen evaluation metrics and comparing it to the benchmark model.
8. **Interpretation and Conclusion:** Discussing the findings, model performance, and providing insights into customer behavior and their response to different offers.

By following this project design, the solution will make use of machine learning techniques to find effective marketing strategies considering the demographic differences among Starbucks rewards mobile app users, ultimately driving sales and customer engagement.