

Starbucks

Domain background:

The project delves into Starbucks' customer behavior analysis, focusing on understanding how individuals engage with their rewards program and respond to various promotional offers. By simulating purchasing decisions, we aim to uncover insights that will guide future marketing strategies. The analysis centers on historical data, including transaction records, customer demographics, and responses to past promotions. By applying machine learning techniques, we will identify distinct customer segments and their preferences. This project draws on academic research in the field of marketing analytics, specifically referencing the work of Kumar et al. (2016), who explored the impact of reward programs on customer behavior. Through our analysis, we aim to answer key questions such as: What types of offers resonate most with each customer segment? How can we optimize the timing and content of our promotions to increase customer satisfaction and loyalty? By understanding these dynamics, Starbucks can design more effective rewards programs, enhance customer engagement, and ultimately, boost sales performance.

Problem statement:

The problem is to identify which groups of people are most responsive to each type of offer (buy-one-get-one, discount, or informational) and determine the best way to present these offers to maximize response rates. The task is to develop a model that can predict customer responsiveness to different offer types, allowing businesses to optimize their promotional campaigns and improve customer engagement.

Datasets and inputs:

- `profile.json`: Contains information about rewards program users, including gender, age, ID, membership date, and income. **(17000 users x 5 fields)**
- `portfolio.json`: Details the offers sent during a 30-day test period, including reward amount, channels used, spending requirement, duration, offer type, and offer ID. **(10 offers x 6 fields)**
- `transcript.json`: Logs user events such as offers received, offers viewed, transactions made, and offers completed, along with associated metadata like person ID, event type, offer ID, amounts spent and rewarded, and time stamps. **(306648 events x 4 fields)**

Solution statement:

Develop a machine learning model to handle such multiclass classification problem. Customers will be classified into one of the three offers based on customer behavior. The model should be able to predict the best offer based on customer behavior. I will be using Jupyter notebook with python version 3.x for this project. I will first try the project locally. If I found that a higher computing power is needed, then I will shift to AWS sagemaker notebooks or google colab.

Benchmark model:

A simple baseline model could be a rule-based system that makes predictions based on average customer behavior. For example, it could assume that all customers have an equal probability of responding to each offer type or that certain customer segments (e.g., based on age or income) are more likely to respond to specific offers.

Evaluation metrics:

- Accuracy: The proportion of correct predictions made by the model.
- F1 Score: The harmonic mean of precision and recall, providing a balanced measure of model performance.

Project design:

The project can be structured as follows:

1. Data Preprocessing: Clean and preprocess the dataset, handling missing values and converting data types as necessary.
2. Exploratory Data Analysis: Analyze the dataset to gain insights into customer behavior, offer effectiveness, and potential patterns or relationships.
3. Feature Engineering: Create new features that capture important aspects of customer behavior, offer characteristics, and time-related factors.
4. Model Selection and Training: Choose appropriate machine learning algorithms, train the models using the preprocessed data, and evaluate their performance using the selected evaluation metrics.
5. Hyperparameter Tuning: Fine-tune the selected models by optimizing their hyperparameters to improve performance.
6. Model Interpretation: Interpret the results and insights gained from the trained models, explaining the factors that influence customer responsiveness.