Optimizing Starbucks App Offers Delivery

Machine Learning Engineer Nanodegree
Capstone Proposal

March 2nd, 2023

M Muchsin

Domain Background

Starbucks offers one of the most sought-after loyalty programs 'Starbucks Rewards' program that offers freebies and discounts to members giving them plenty of reasons to choose Starbucks over other players. 'Starbucks Rewards' program represents a significant portion of the coffee chain's recent fiscal growth. Starbucks has reported an increased revenue of $2.65 billion, attributing their rewards program for most of the increase. Over the last two years, membership has grown more than 25%, loyal customers use Starbucks' membership program (16 million members) for about 40% of sales at the company's US stores. Revenue rose 4.6% to $6.31 billion during the quarter from the previous year.

However, with so many offers available, it can be difficult for Starbucks to optimize its offers to ensure that they are attractive to customers while also being profitable for the company. Additionally, Starbucks must consider the preferences and behavior of individual customers when creating these offers, as different customers may respond differently to different types of promotions.

To address these challenges, Starbucks seeks to optimize its app offers using data-driven approaches. By analyzing customer data and behavior, Starbucks hopes to create personalized offers that are more likely to be accepted and lead to increased customer engagement and loyalty. Additionally, by optimizing these offers, Starbucks can improve its profitability and maintain its position as a leader in the competitive coffee market.

Problem Statement

The problem at hand is how to effectively segment customers using demographic and behavioral data to create personalized offers that maximize the conversion rate. By identifying subgroups of customers with similar characteristics, businesses can develop targeted marketing campaigns that are more likely to resonate with each group. This involves analyzing customer data to identify patterns and trends in their behavior, preferences, and demographics. Once these patterns are identified, businesses can develop personalized offers that are tailored to the needs and preferences of each customer segment. Machine learning algorithms can be used to predict which offers are most likely to be successful for each segment. By continually refining and improving the offer selection process, businesses can maximize the conversion rate and drive revenue growth.

Datasets and Inputs

The datasets used in this project are 3 JSON files with the following details:

1. **Profile.json**

   Rewards 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)

2. **Portfolio.json**

   Offers sent during the 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 a reward
   - duration: (numeric) time for the offer to be open, in days
   - offer_type: (string) bogo, discount, informational
   - id: (string/hash)

3. **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 the start of the test

Solution Statement

An ensemble model will be used in this project. Ensemble learning is a technique that combines multiple base estimators to produce better predictions than any single estimator. The project will use AutoGluon, an open-source framework that automates the process of creating and tuning ensemble models.

Benchmark Model

Decision tree as a simple model will be used for the benchmark model. It is a simple and interpretable model that can handle both categorical and numerical features. It also provides a baseline for comparing more complex models such as random forest and gradient boosting.

Evaluation Metrics

There are some options for the metrics. If the objective is to maximize customer satisfaction by sending relevant offers, precision is better than recall. If the objective is to increase customer engagement by sending more offers, recall is a better option than precision. In this project, F1 score will be used as the evaluation metric. It is a good metric when the objective is to balance between precision and recall, especially when the classes are imbalanced.

Project Design

1. Exploratory Data Analysis
2. Data Cleaning
3. Feature Engineering
4. Model Training
5. Hyperparameter Tuning

Reference

https://www.zinrelo.com/loyalty-rewards-case-study-new-starbucks-rewards-program.html

https://www.forbes.com/sites/bernardmarr/2018/05/28/starbucks-using-big-data-analytics-and-artificial-intelligence-to-boost-performance/?sh=76f1a1c065cd

https://blog.hubspot.com/service/customer-segmentation

https://auto.gluon.ai/stable/_modules/autogluon/core/models/ensemble/weighted_ensemble_model.html

https://www.ibm.com/topics/decision-trees

https://medium.com/analytics-vidhya/confusion-matrix-accuracy-precision-recall-f1-score-ade299cf63cd