Starbucks Capstone Challenge Proposal

Domain background

As one of the biggest coffee giants in the world, Starbucks operates more than 31,000 stores worldwide, in 2017, Starbucks launched its most advanced AI-driven initiative "Deep Brew". The brand's custom-made recommendation platform was built to reach customers across multiple channels, including the Starbucks ordering app. With the platform, it delivers highly customized offers to the almost 19 members of its My Starbucks Rewards loyalty program. It helps Starbucks to deploy individualized offers across channels by automating offer assembly and management, reward fulfillment, and KPI measurement and tracking, at enterprise scale.

The industry-leading Starbucks Rewards Program has continued to flourish since its introduction in 2007. According to their website, "Membership has grown more than 25% over the past two years alone, climbing to 16 million active members as of December 2018, a 14% increase over the prior year. Starbucks Rewards transactions accounted for 40% of tender in U.S. company-operated stores in the same time frame." Obviously, it becomes more and more important to leverage the AI technology to enhance the customer experience.

Problem statement

Since it's important to provide customized experience from ordering to offer on loyalty app, the problem becomes to how to correctly and precisely locate the target customer group. That is, the task is to decide what the best-personalized offer is to send to users so that the conversion rate can be maximized. Also, it's critical to know which groups of people are most responsive to each type of offer, and how best to present each type of offer.

Datasets and inputs

There are three data files provided in this project including demographic, offer and transaction data from the Starbucks rewards mobile app.

The profile dataset contains Rewards program user's information which gives data like gender, age, income and time when the customer became a member. The portfolio dataset contains the list of all categories of offers sent to the customer during 30-day test period. There are three types of offers that can be sent: buy-one-get-one (BOGO), discount, and informational. And the last dataset is transcript which gives event showing different actions (e.g., offer received, offer viewed) and amount of the transaction spent on the offer.

Solution statement

The plan of this project is to build a machine learning model to decide which is the best type of offer to send to each customer. The dataset will be separated into three types of offer and fit the data into three supervised classification models. Using this way, the model will predict whether the offer will be responded by customer or not when sent to them. Also, by investigating the feature importance of the model, it can help answer the question that what factors mainly affect people to make the decision and finally complete the transaction. Therefore, with the solution above, it'll be easier and more accurate to identify which groups of people are most responsive to each type of offer, and how best to present each type of offer.

Benchmark model

It’s always a good practice to set a benchmark while building further fancy machine learning models. Therefore, in this project, it will use the current conversion rate of each type of offer as the benchmark. For BOGO and discount offer, the current conversion rate will be the percentage of offers which completed among all offers viewed after received. And for informational offer, since there’s no explicit offer completed, here consider the transactions completed within 1 day since the informational offer was received. So, the overall current conversion rate is 52% and for each group the current conversion rate is as following.

In this project, first the simple decision tree classifier and random forest algorithm will be selected to solve the problem. Then, the hyperparameters tuning will help to find out the optimized parameters used in the final models. In the end, it will compare the model with best performance to the benchmark to decide whether the model is capable to solve the problem stated earlier in this project.

Evaluation metrics

Since the project is building classification model, here choose both accuracy and F1 score as the model evaluation metric. The reason of choosing both metrics is sometimes when the dataset is imbalanced, the accuracy only couldn't objectively show how the model is performing on the dataset, while F1 score provides a better sense of model performance compared to purely accuracy as takes both false positives and false negatives in the calculation. With an imbalanced class distribution, F1 may be more useful than accuracy.

Also, since the F1 score is based on the harmonic mean of precision and recall and focuses on positive cases. For the Starbucks app here, it would be fine as we would prioritize more on whether offers are effective, and less focus on why offers are ineffective.

Project design

The project is designed with the following steps:

Prepare and clean data -- combine transaction, demographic and offer data. Understand the connection between columns and dataset. Try to get the useful information from data as much as possible. A good understanding on dataset is essential to next step of exploration.

Data exploration -- In order to analyze the problem better in next sections, first need to explore the datasets which includes checking the missing value, visualizing the data distribution, etc. In that way, we can have a better understanding on how the dataset looks like and how to select the important features to support the model implementation.

Data preprocessing -- In order to find out what mainly affect the finish of the transaction by sending the offer, in the data processing process, also need to process the data to merge the events of each specific offer sent so as to find out which offer were received, viewed and finally completed with a transaction.

Feature engineering -- After basic processing, the next step will look if there are any columns that can be used to create new features. For example, generating a new column for length of customer's membership, the count of offer received for each user, calculate the time lap between offers, etc.

Building model – after pre-processing and feature engineering, next step is to build the model using response flag generated in previous steps to predict whether the customer will respond to the offer or not. Here we will choose the basic tree model as a baseline which will help explain the feature importance better so that we can get some insight into what factors affect customer's behavior most. Meanwhile, I'll also choose random forest as an alternate model to compare the baseline model is as an improvement over simple ensemble bagging of decision trees, in order to drive towards a high accuracy in training the model.

Model tuning – Compare the model using metrics selected above and tune the parameters of initial model using GridSearch method to get higher performance.

Conclusion and further improvement – compare the final selected model to benchmark to see if the solution provide a better personalized offer. Also, review the built process and see if there's any opportunities to enhance the model in the future.