

# Starbucks Capstone Challenge – Proposal

## Domain background

The domain of this project is the exploration of customer behavior, particularly the customer behavior when presented with advertisement or product offers. More specifically, the behavior of customers within the Starbucks rewards app. It is to be investigated how Starbucks should present clients with offers to maximize the effectiveness of the offer without cannibalizing existing revenue. The offers can consist of simple informational advertisement or an actual offer such as discounts. The goal for Starbucks is to place the offers in their app in a way that maximizes their effectiveness, i.e. presenting the right offer to the right customer. For every offer sent there are 4 possible outcomes based on whether the offer is viewed and completed. These outcomes can be ranked from most desirable to least desirable (1 – most desirable to 4 – least desirable):

|            | Completed | Not Completed |
|------------|-----------|---------------|
| Viewed     | 1         | 2             |
| Not Viewed | 4         | 3             |

## Problem Statement

In order to make good recommendations to customers it is helpful to know what the expected outcome of an offer presented to a specific customer will be. The problem statement is therefore to build a model that can predict whether a customer will view and complete an offer based on the type of offer and the characteristics of the customer.

On the basis of this model a simple recommendation function can be built that determines the offer to send to a customer.

## Datasets and Inputs

There are three datasets given to solve this problem. Firstly, there is a dataset with the portfolio of offers. It consists of ten entries that define the type of offer, its duration, its difficulty, its reward and the channels used by this offer. Each offer has a unique ID.

The second dataset describes the customers. It provides information on the age, gender and income of each customer and holds seventeen thousand entries.

The third dataset is a record of past transactions. This dataset holds over 300.000 entries about received offers, viewed offers, completed offers made by customers and the corresponding offer ID and reward of the completed offers. Additionally, transactions of customers are recorded with

the amount of each transaction. The dataset further records the time at which the transaction occurred.

From these datasets an input dataset can be build. This input dataset should include the profile information of a customer, the average transaction volume of a customer per week over the recorded time-horizon and the type of offer that will be sent to the customer.

The output of the model will show into which category of the viewed/completed matrix the offer for a customer is most likely to fall.

## **Solution Statement**

To solve this problem a machine learning model will be trained on the dataset. As the data provided insight on the actual outcome, this model will be a supervised learning model. A number of different techniques will be tested such as a neural network, an XGB Boost model or a Naïve Bayes model.

## **Benchmark Model**

As a benchmark a simple model will be trained on the dataset. An example for such a model could be Sagemaker's built-in Linear Learner.

## **Evaluation Metrics**

For the evaluation metrics the standard metrics used for machine learning models such as accuracy, recall, precision and the F-Score will be used.

## **Project Design**

In a first step the datasets will be cleaned. This means dropping any customer profiles that are incomplete and the transactions recorded from these incomplete customer profiles. Next the transaction volume per customer per week will be added as additional information to the customer profiles. Lastly, a dataset that records all the offers sent to customers and their outcomes (i.e. whether they were viewed and whether they were completed) will be constructed. For every offer sent the personal information of the customer will be added to the offer dataset to construct the input dataset explained above.

After the dataset has been split into test and training set, the Benchmark model will be trained to establish the Benchmark.

Next several different Machine Learning models will be trained and tested to see which model performs best based on the evaluation metrics.

Lastly, once the best model has been picked, a sample recommendation function will be introduced. This recommendation sample will take the customer ID as input and then predict the outcomes for the available offers for this customers. Once the outcomes are known a predetermined ranking will identify which offer is most suitable for the customer and this offer will be returned.