

Starbucks Capstone Project Proposal

Udacity Machine Learning Engineer

Nanodegree

Cole Lineberry
June 28, 2022

Table of Contents

Domain Background	1
Problem Statement	1
Datasets and inputs	2
Solution Statement	2
Benchmark Model	2
Evaluation Metrics	2
Project Design	2

Domain Background

Starbucks is a world wide coffee maker and one of the Fortune 500 companies. They have thousands of customers who buy their coffee daily. These customers have the option to use a mobile application to purchase coffee. The company has gathered lots of customer data and demographics from these mobile apps. The marketing team at Starbucks has been offering some buy one get one free promotions (BOGO). These are hard to predict if they will be successful. Not all users receive an offer every week and not all users receive the same offer.

Problem Statement

Starbucks would like to know which offers to send to which demographic groups based on previous responses to offers. They would also like to be able to predict which offers will most likely be completed by these groups.

Datasets and inputs

The Dataset contains simulated data that mimics customer behavior on the Starbucks rewards mobile app. Once every few days, Starbucks sends out an offer to users of the mobile app. An offer can be an ad for a drink or an actual offer for a discount or BOGO. Some users might not receive any offer during certain weeks.

The data is contained in three files:

- portfolio.json - containing offer ids and meta data about each offer (duration, type, etc.)
- profile.json - demographic data for each customer
- transcript.json - records for transactions, offers received, offers viewed, and offers completed

Solution Statement

I will be building a machine learning model to predict the customer response to an offer. This will be based on the customers previous responses, taking into account customer demographics. This will include doing some data analysis at the beginning of the project to best set up the data for demographics vs customer response. Then I will build a machine learning model based on training data and compare that to a test set of data.

Benchmark Model

I will use a KNearestNeighbors classifier to create a benchmark for the machine learning model. KNearestNeighbor is a good algorithm for this as it will create a nice binary classification that will be easy to test against. This will provide a way to tell if my predictor is performing at or better than the benchmark algorithm.

Evaluation Metrics

The KNearestNeighbor algorithm will provide a good way to classify the data as either a positive response, no response, or incomplete response. This will provide an F1 score which will be a number between 1 and 0 with anything closer to 1 being the most likely response. F1 provides the balance between precision and recall and is calculated with $2((\text{precision recall})/(\text{precision} + \text{recall}))$

Project Design

This project will be designed and built within a jupyter notebook. The project will be laid out in the following steps,

1. Data preparation: Obtain the data sets and remove any values that are not necessary for later use.

2. Data Exploration: Create visualizations on the data to get a better picture of the overall data set. Perform any tuning or refining to better figure out what the best predictive model for this type of data will be.
3. Data transformation: Make the necessary data wrangling operations to be able to feed the training and test data into a machine learning algorithm.
4. Train model: Use the values from the training set to train the model
5. Predict outcomes: Use the model to predict outcomes based on test data.
6. Conclusion: Compare the benchmark to the predicted values and then summarize the findings.