

# CAPSTONE PROJECT PROPOSAL

Machine Learning Engineer Nanodegree

EDIDIONG ALIGBE

March 21, 2023.

## **Domain Background**

Starbucks is an American chain of coffeehouse, located in over 80 countries of the world and has its headquarters in Seattle, Washington, United States of America. It is known for its high-quality coffee and espresso-based drinks like Caffè Latte, Americano, and Cappuccino. It also offers hot chocolate, cold beverages, teas, salads, pastries and sandwiches.

Starbucks offers promotions designed to attract new customers and encourage existing customers to make repeat purchases, building brand loyalty. These promotions could be in form of loyalty programs, buy one, get one free (BOGO), gift cards, etc. These offers come through emails, the web, mobile app, and social media. The company keeps a record of customers' interactions with these promotions and the data is useful to serve the customer better based on their activities and preferences

Machine Learning is an area of Artificial Intelligence that allows software applications to learn from data and predict outcomes without being explicitly programmed to do so. With more data, they learn to be more accurate and improve their performance.

I am interested in using Machine Learning and the data from customers' interactions with promotions to help Starbucks to run their promotions successfully.

## **Problem statement**

The aim of sending promotions and offers to customers is to encourage customers to make purchases. It will not be an intelligent move to send the same promotion to all the customers at once, this may result in promotions that do not yield the expected results. The goal of this project is to make use of the data from Starbucks to determine what promotion to send to a demography of customers that will result in higher customer purchases.

## **Dataset and inputs**

The dataset for this project contains simulated data that mimics customer behaviour on the Starbucks mobile app. These data comes in three json files:

1. profile.json: Contains demographic data for each customer. It is made up of 17,000 customers with 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: Contains data on offers sent during 30-day test period. It is made up of 10 offers with 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 reward
  - duration: (numeric) time for offer to be open, in days
  - offer\_type: (string) bogo, discount, informational
  - id: (string/hash)
3. transcript.json: Contains records for customer transactions, offers received, offers viewed, and offers completed. It is made up of 306648 transactions with 4 fields -
  - person: (string/hash)
  - event: (string) offer received, offer
  - 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 start of test

## **Solution statement**

Proposed solution for developing a machine learning model that will determine which promotion to send to a demography of customers will take the steps below:

1. Data collection.
2. Exploratory data analysis and data visualization.
3. Model training.
4. Model evaluation.

## **Benchmark model**

This project will make use of AWS AutoGluon to easily achieve a high performing model. AutoGluon will train a variety of supervised classification models and the best performing model will be used as a benchmark model.

## **Evaluation metrics**

This project will make use of machine learning classification models to provide solution. Evaluation metrics in this project will be:

1. Accuracy
2. Precision
3. Recall.

## **Project design**

**Objective:** Develop a machine learning model that will determine what offer to send to a demography of customers.

**Resources needed:**

1. Data
2. Software: AWS Sagemaker, AutoGluon, Pandas, Matplotlib, Scikit-Learn
3. Hardware: Computer system with sufficient processing power and memory.

**Task:**

1. Data collection and cleaning
2. Exploratory data analysis - Perform descriptive statistics to understand the distribution of variables, and conduct visualizations to identify patterns and trends in the data.
3. Feature engineering - create new features that may improve the predictive power of the model, and encode categorical variables.
4. Model training and evaluation using AWS AutoGluon on train and test datasets.
5. Model Tuning and validation - fine-tune hyperparameters to improve performance and check for overfitting.
6. Generate project report and submit.

**Timeline:** The project timeline is 1 week.