Project Proposal: Starbucks Capstone Challenge

Jared Gan

Udacity – AWS Machine Learning Nanodegree 2023

Domain Background:

Starbucks is one of the world's largest coffeehouse chains that has its presence in 80+ countries worldwide. Starbucks is known for its customer-focused strategy and providing an excellent in-store experience to customers. In addition to that, Starbucks has a mobile app that customers can use to pay for their orders, earn rewards, and avail of special offers. To ensure that Starbucks is sending the right offers to the right customers, it is crucial to analyze customer behavior on the mobile app. Starbucks is keen on identifying which demographic groups respond best to which offer types, and that's where this challenge comes in.

Problem Statement:

The challenge is to analyze the customer behavior on the Starbucks rewards mobile app and identify which demographic groups respond best to which offer types. Starbucks sends out different types of offers to its customers, including Buy One Get One (BOGO), discounts, and informational offers. Not all customers receive the same offer, and some customers may not receive any offer during certain weeks. The goal is to build a machine learning to predict customer behavior upon receiving offers from Starbucks. Thus, enabling Starbucks to tailor its marketing campaigns accordingly.

Proposed Solution:

To address this challenge, I propose to perform exploratory data analysis (EDA) on the provided dataset to understand the relationship between customer demographics, offer types, and customer behavior on the mobile app. The next step would be to preprocess and clean the data, ensuring data integrity and addressing any missing values. Once the data is clean, I will use machine learning techniques to build a predictive model that can determine which customers respond best to offers.

Deliverables:

The deliverables of this project would be a well-documented codebase that includes data preprocessing, EDA, and a predictive model. The model should be able to predict which customers are more likely to make a purchase upon receiving offers. The project will be presented in the form of a technical report that documents the entire project's scope, methodology, results, and conclusions. The report will also include visualizations to help explain the findings and a detailed discussion of the limitations and future scope of the project.

Evaluation Metric:

We will be using the F1 score as our primary evaluation metric for this project. The F1 score is a measure of a model's accuracy that considers both precision and recall. It is a useful metric when dealing with imbalanced datasets, which is common in many real-world scenarios. The F1 score is calculated as the harmonic mean of precision and recall, and it ranges from 0 to 1. A score of 1 indicates perfect precision and recall, while a score of 0 indicates that the model has no predictive power. By using the F1 score as our primary metric, we can ensure that our model performs well in terms of both precision and recall, which are both important for our project's goals.

Project Design:

1. Create an Amazon SageMaker notebook instance and set up the required environment.
2. Import the necessary libraries such as pandas, numpy, seaborn, matplotlib, and sagemaker.
3. Load the Starbucks dataset into the notebook instance and explore it to gain a better understanding of the data.
4. Preprocess the data by performing tasks such as data cleaning, feature engineering, and feature scaling.
5. Split the data into training, validation, and test sets.
6. Train a machine learning model using one of the available algorithms such as XGBoost, Random Forest, or Deep Learning.
7. Optimize the model's hyperparameters using techniques such as grid search or Bayesian optimization.
8. Evaluate the model's performance on the test set and compare it with the performance of other models.

Conclusion:

This project aims to analyze the customer behavior on the Starbucks rewards mobile app and identify which customers respond best to offes. The proposed solution involves data preprocessing, exploratory data analysis, and machine learning model development. The project's success will be measured by the accuracy and reliability of the predictive model developed and the insights it can provide Starbucks regarding their mobile app's marketing campaigns.