Starbucks Capstone Proposal

Abdul Tamimi

15/12/2021

Domain Background

Starbucks is just like any powerful company where its business is based on customer interaction. Starbucks owns stores worldwide and it has succeeded to achieve that feat by focusing on customer behaviour relentlessly.

One of the greatest conveniences given to customers was the ability to interact within the Starbucks mobile app. Customers can receive, view, complete offers through this and other channels. Customers can even collect rewards and various offer types through the introduction of the Starbucks app. The introduction of the app has also on the other hand introduced concerns that if solved could enhance customer satisfaction and bring more revenues, One of those challenges is to monitor how customers make purchasing decisions. Each customer has different traits that affect their purchasing behaviour and pattern, for that the challenge is to identify how different groups of people respond to offers and rewards sent via the Starbucks app. The challenge is to create a machine learning model to analyze and predict how customers will respond when receiving an offer.

Problem Statement

In this project, we are trying to solve one main problem and that is to see which group of people are most responsive to each offer type, and how each demographic group responds to offer types. Each customer has hidden traits that influence their purchasing patterns and accordingly respond to offers differently. To analyze customers purchasing decisions will build machine learning models to predict how a customer will respond to an offer.

Datasets and Inputs

The dataset used in the project is provided by Starbucks. There are 3 datasets that each contain stimulated data that mimics the behavior of a customer while using Starbucks mobile app.

Each dataset contains different information to monitor how purchasing decisions influence promotional offers. The 3 datasets are: profile.json, portfolio.json and transcript.json. The profile dataset includes information about gender, age and income of the customer and date in which the customer became a member on. The portfolio dataset includes more information on the offer itself. A reward column describing an amount in which the customer receives when completing an offer. A difficulty column describing the amount needed for a customer to spend to receive reward. The dataset also contains offer types and duration in days that shows the time in which these offers are open for. The last dataset, transcript.json, contains information on the event type and their corresponding value.

Solution Statement

To be able to understand purchasing patterns will build a model to predict how a customer reacts when receiving an offer. That is if a customer receives an offer what is the percentage that customer will complete that offer. Explatornayr data analysis will be our guide to build different models and algorithms and see which model(s) will work best for our classification problem.

To find out whether the customer will respond to an offer, will be using the algorithms like random forest classifier, logistic regression, naive bayes, kneighbours classifier and decision tree.

Evaluation Metrics

In this project, I will evaluate the machine learning algorithms using a confusion matrix and classification report. Since this is a classification problem, classification reports will help summarize the performance of the problem. Confusion matrix also helps enhance the effectiveness and the performance and produce visualizations to better the problem. Confusion matrix includes several important metrics that will help measure the performance of the model: precision, recall, f1 score and accuracy.

Project Design

The workflow that will follow to address the problem is as follows. First will load and clean the datasets. Secondly, apply data preprocessing to help make the datasets better for visualizing and analysis. Will apply feature engineering where some features will be needed to be scaled and normalized to skew the targets for modelling. After normalising, will merge the three datasets to create train and test datasets. After that, we will fit, train and predict our model using the algorithms mentioned above. Finally, measuring the performance of our models by using evaluation metrics.