Capstone Proposal

Machine Learning Engineer Nanodegree

Domain Background

Retail businesses give offers and promotions to their customers in order to retain their customer base, influencing them to spend more or even to try different products which are not in their typical basket.

Starbucks is an American multinational chain of coffeehouses and also the world's largest coffeehouse chain. As of September 2020, the company had 32,660 stores in 83 countries, including 16,637 company-operated stores and majority of which are in North America.

According to this case study, the Starbuks mainly promotes their products through rewards to its customers through their mobile app. It led their customers to purchase more or stay in touch with the brand for a longer time. Starbucks sends out offers to users of the mobile app once every few days.

An offer can be any of the following;

1. an advertisement for a drink or
2. an actual offer such as a discount or
3. BOGO (buy one get one free)

Some users might not receive any offers during certain weeks and offers will be different to every customer, and offers have validity periods as well.

Problem Statement

Customers receive different types of offers. It affects their purchasing pattern and the relationship with the brand as well. Customers from different demographic groups may respond to offers differently.

This study attempts to analyse the following problems;

1. which offers or offer types receive better responses from the customers
2. Whether there are any specific groups which responds well to the offers
3. How to recommend offers to the customers based on their past interaction with other offers

Exploratory data analysis, Unsupervised learning methods and Recommendation system designs would be a suitable approach to respond to the above problems.

Datasets and Inputs

The data is contained in three files:

- portfolio.json - contains offer id 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

Above datasets captured 17000 unique users' behaviours on offers for a period of a month. Here is the schema and explanation of each variable in the files:

portfolio.json

- id (string) - offer id
- offer_type (string) - type of offer, i.e. BOGO, discount, informational
- difficulty (int) - minimum required spend to complete an offer
- reward (int) - reward given for completing an offer
- duration (int) - time for offer to be open, in days
- channels (list of strings)

profile.json

- age (int) - age of the customer
- became_member_on (int) - date when customer created an app account
- gender (str) - gender of the customer (note some entries contain 'O' for other rather than M or F)
- id (str) - customer id
- income (float) - customer's income

transcript.json

- event (str) - record description (ie transaction, offer received, offer viewed, etc.)
- person (str) - customer id
- time (int) - time in hours since the start of the test. The data begins at time t=0
- value - (dict of strings) - either an offer id or transaction amount depending on the record

Solution Statement

Exploratory data analysis (EDA) would reveal key information such as which offers were more popular or utilized more and which offers influenced customers to buy or spend more at the outlets. Also EDA would help to understand democratic background of the customers of Starbucks.

Identifying the best type of offer for an individual or a group of individuals who share similar characteristics is critical.

- Unsupervised learning methods will help to group customers from different demographic groups and their response to each offers that they received

- ○ Unsupervised algorithms such as KMeans or Agglomerative Clustering will be employed to spot groups of individuals who respond significantly differently to the offers to understand their democratic and other features.
- Recommendation system would be the right approach for identifying appropriate offers to the customers at individual level
  - ○ Collaborative Filtering algorithms could be employed to recommend best offer / offers that will be responded to positively by the customers

Benchmark Model

Average completion rate of an offer or offer type would be a good benchmark to compare the performance of the model or methods that are proposed.

If a method or model outperforms the average offer completion rate that can be considered a better model than the existing model or method.

Evaluation Metrics

As the Starbucks case is defined as a recommendation problem, recommending an offer that could be utilised by a customer should give higher scores to the model. Therefore the deviation between actual offer completion rate and predicted score for the offer type at offer level should be measured and a RMSE measure will give a better picture about the total deviation matrix. The model could also be refined to minimize the deviation.

Project Design

Identifying the right offer type at an individual level or group level is the objective of this study.

Following steps are proposed to approach the above problem:

1. Understanding the business use case of Starbucks
2. Understanding the dataset
3. Cleaning, preprocessing and preparing the dataset
4. Exploratory Data Analysis
5. Advanced analysis
6. Evaluating the models and hypothesis
7. Communicating the results