Machine Learning Engineer Nanodegree, Udacity

Starbuck Campaign Challenge Proposal

Wasurat Soontronchai

16 June 2020

Proposal

1.) Domain Background

Starbuck is worldwide coffee company from America. Currently the company have more than 24,000 stores across more than 75 countries.
On every day, Starbucks promote their offer to customer via mobile application platform.

Their offer can be advertiment of either a discount, BOGO (Buy One Get One Free) or do not receive any offer in some period. Each offers has period of acceptance for customer before its expire. For example, 7 days of offer validity.

So, how these data can be valuable to the company ?
If company can send the offers to the target customer, it would help them to generate more benefit from higher offer accepted rate !

This come to my problem statement for this capstone project. Can we predict whick kind of customer tend to use or accept the certain type of offer ?

I select this problem because Starbuck is my every morning coffee. Discount and BOGO also are campaigns where I normally receive from email and sometimes in message this seem to be more realistic problem for me as a premium member of the Starbuck application.

2.) Problem Statement

As mentioned in domain background, the problem statement is to predict wheter customer will accept to an offer from given demographics informations and type of campaign (BOGO, discount etc.).
The problem will be **Binary Classification**

> Postive class : customer accept the offer (1)
> Negative class : customer do not accept the offer (0)

3.) Datasets and inputs

Total data available come in .json format 3 files:

1. portfolio.json : offer ids and detail of offer (durations, type of offer, etc.)

id (str) : offer id
offer_type (str) : BOGO, discount, intormational
difficulty (int) : minimum effort spent to complete offer
reward (int) : reward for completing offer
duration (int) : limit time for offer before expire (days)

channels (str list) : channel of communication in offer

2. profile.json : demograhic of customer

age (int) : age of customer

became_member_on (int) : start date of account creation in application

gender (str) : gender of customer

id (std) : customer id

income (float) : estimate customer's income

3. transcript.json: transactions record, offer recieved , viewed, offer completed

event (str) : transacton description

person (str) : customer id (match on profile.json)

time (str) : hours of start the campaign test

value (str dict) : amount of offer id or transaction

4.) Solution Statement

Firstly, combine the 3 files to make it easier for analysis in later stage. Table will have all provided information about customer (demographic), offer details, transcation data and the last one is target prediction which is offer acceptance from customer (0 or 1)

Then use all feature execpt offer acceptance from customer as input features of binary classifiers model.

5.) Benchmark Model

To benchmark the model performanace, I would set the lowest acceptance criteria to beat the naive model performance on selected evaluation metrics which is assumes all offers were successful.

6.) Evaluation Metrics

As the problem is binary classification, I choose

- Accuracy : To measure over model correction
- F1-score : To consider both precision and recall for model accuracy.

7.) Project Design

Start with fundamental process of data science project as below.

- Cleaning the data (Dealing with missing value, outliers, data type etc.)
- Perform data exploratory analysis to get more insight from the data.
- Feature Engineering (Grouping, drop out unrelated columns)
- Model development (Trial all binary classifier)
- Model Evaluation
- Model Tuning (Base on selected classifier)

All of these process shall be iterable process.