Proposal

LarebSyed

Domain Background

The Starbucks Capstone Challenge is an classic example of a sales problem where we want to make an offer to the customers that they can't refuse. We can solve the problem by looking at the past behavior of the customer and predict what items or offers the customer will be most interested in. By surfacing the right offers to the customer we can generate more sales. We have with us the information of past transactions of customers and related offers which were made to the customers.

Research and Citation

- https://www.sciencedirect.com/science/article/abs/pii/S0957417412006148
- https://martech.org/machine-learning-for-next-best-offers/
- https://towardsdatascience.com/implementing-a-profitable-promotional-strategy-for-starbucks-with-machine-learning-part-1-2f25ec9ae00c
-

Problem Statement

Problem

Generate a system that will predict the type of offer to make to a customer given the customer demographics and transaction history

Solution

- Analyze and clean the starbucks historical transactional data
- Use machine learning to train a model that can predict the offers to send

Datasets and Inputs

We have three data artifact to use in designing the system. We have 10 different types of offer and we will classifying all of our customers in these 10 classes. The number of transaction are 306534 and 17000 users demographics

portfolio.json

This contains all the available offers we can make to customer, it contains following characteristics

- id (string) - offer id

- $offer_{type}$ (string) - type of offer ie 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

This contains the all the information related to customers including their demographics. This will be used in training the model for offer predictions.

- age (int) - age of the customer
- $became_{memberon}$ (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

This contains list of all the orders made by the customers. This will also be used in conjunction with profile.json to train our ML model

- event (str) - record description (ie transaction, offer received, offer viewed, etc.)
- person (str) - customer id
- time (int) - time in hours since start of 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

Solution will comprise of following steps

- Analyze and clean the starbucks historical transactional data
- Use machine learning to train a model that can predict the offers to send
- use the model to predict future customer offers

Benchmark Model

We will benchmark our model with logistic regression model.

Evaluation Metrics

We will use F_β score with $\beta=2$ to give more weightage to recall as according to our use case it is better to show a wrong offer to a customer instead of not showing the right offer to a customer. But still we don't want to overwhelm a customer with wrong offer thus be are use F score.

Project Design

To design a model for this problem we will be using AdaBoost model, we will first analyze the data to clean up unwanted features and nil values. Then we will process the data to handle categorical data and also perform data standardization where needed.

Then well use the processed data to train and evaluate of our benchmark model to compare with our chosen model.

Finally, we will train and tune our AdaBoost model to achieve our goal As a fallback we can also use XGBoost model if we don't get desired results from AdaBoost