Machine Learning Engineer Nanodegree Program

Capstone Project Proposal

Starbucks Capstone

Devesh Katiyar
Feb-2020

Domain Background

Machine learning has become an increasingly important part of IT today and usage of ML is increasing daily in almost every domain. However, the power of ML is not utilised to its fullest yet and can be useful for society and market in unimaginable ways. ML can also be useful for various companies in improving their products, sales to offer better customer service and earning a higher growth rate.
STARBUCKS is one of the flagship Worldwide companies which has been established since Mar-1971 and worldwide coffeehouse chain, and has a tremendous database of users. Starbucks offers their free app to make orders online, predict the waiting time and offer better service. That's why analysis on app usage is more crucial to leverage the business and understand the customer's behavior.

Problem Statement

As Mentioned in **STARBUCKS Capstone Challenge,** analyze the Starbucks Customer dataset and build a **model that can make better offers to customers**. Our goal is to analyze the dataset made available by Starbucks about the app usage and offers/orders made by the customer to make a model that can make better offers to the customers so that they redeem the offer made by the model which will finally increase the sales of the Starbucks and help them to reach more new customers.

Datasets and inputs

The data is contained in three files:

- portfolio.json - containing offer ids 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

portfolio.json

- 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

- 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

My strategy is to develop a Machine Learning model to predict which is the type of offer for each customer, such that the offer proposed to the customer turns into a sale. I will develop a model for each offer type and then combine the results to have the best action for each user.

Benchmark Model

We will use different models to get the best out of our dataset.

This will include Decision Tree Classifier, Adaboost Classifier and RandomForest Classifier and the model which will show the best performance will be our final model for the implementation.

Evaluation Metrics

We will use accuracy score and F-score as the evaluation metrics of our model.

F-score is a metric that is based on the concept of Precision and Recall, hence it is better for our model evaluation.

Project Design

1. Data Preparation
2. Data Exploration
3. Data Preprocessing
4. Developing the model
5. Choosing the best model and Hyperparameter tuning
6. Measuring the performance