

Machine Learning Engineer - Capstone Project Proposal

Joao Felipe Guedes da Silva
guedes.joaofelipe@poli.ufrj.br

April 7, 2020

1 Project Definition

1.1 Problem Overview

This project aims at building data-driven customer profiles from Starbucks and proposes a model to recommend marketing channels that sends out discount offers based on these customers profiles.

1.2 Problem Statement

Customer engagement is one of the main problems to be tackled nowadays on a digital-based scenario. From making a customer being aware of a brand all the way to making him advocate for it, different levels of difficulties are posed and each of them need to be overcome. The steps a customer needs to is formulated in the 5A's framework [1] which is illustrated below.



One way to enhance customer engagement once he is aware of the brand is to offer him discounts through various marketing channels. However, digital users have different consumption profiles and figuring out what are those profiles may determine which offer approach to take.

In this sense, this project proposes creating customers profiles based on data through clustering techniques. Once these profiles are created, a marketing channel recommender model is developed so as to try to maximize offer view and, eventually, offer consumption.

1.3 Dataset and Techniques

The dataset to be used contains simulated data from the Starbucks rewards mobile app. Frequently, Starbucks sends its users discount offers to be used when a challenge is completed. Demographic informations like age and genders are provided, as well as characteristics of the available offers (difficulty, reward amount, etc.). Ultimately, transaction information for a whole year period is provided indicating which offer was made, which channel was used and whether or not the user completed the offer (thus gaining the reward).

In order to build customer profiles, we propose using a non-supervised clustering technique. Since most users' profile data are numerical, algorithms like K-means [2] can be used to map those users into profiles. The clustering effectiveness shall be evaluated by running a post-training analysis of the clusters with business-oriented variables and through dissimilarities measures.

Once the profiles have been established, a recommender system model is to be created in order to suggest which marketing channel needs to be used so as to maximize the probability of a user viewing an offer. In particular, given the problem statement, we shall develop a multi-armed bandit recommender [3] in order to make a single channel recommendation.

1.4 Figures of Merit

The recommender system shall be evaluated using consumption-based figures of merit such as precision and recall [4]. Also, in terms of business evaluation, offers shall be evaluated in terms of users' advocacy. In other words, once a channel and an offer have been made, we want to understand how likely is this user to consumer the offer and continue making purchases on the app.

1.5 Project Design

All project will be developed in the Amazon Web Service (AWS) environment. The dataset will be stored in a Simple Storage Service (S3) unit and the models will be developed using the Sagemaker service.

After creating customers' data-drive profiles and the channel's recommender system, an endpoint will be created in order to provide a production-ready interface for the project.

References

- [1] P. Kotler, *Marketing 4.0 - Moving from traditional to digital*. Hoboken, 2017.
- [2] J. A. Hartigan and M. A. Wong, "A k-means clustering algorithm," *JSTOR: Applied Statistics*, vol. 28, no. 1, pp. 100–108, 1979.
- [3] J. Vermorel and M. Mohri, "Multi-armed bandit algorithms and empirical evaluation," in *Proceedings of the 16th European Conference on Machine Learning*, ser. ECML'05. Berlin, Heidelberg: Springer-Verlag, 2005, p. 437–448. [Online]. Available: https://doi.org/10.1007/11564096_42
- [4] G. Shani and A. Gunawardana, "Evaluating recommendation systems." in *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, Eds. Springer, 2011, pp. 257–297. [Online]. Available: <http://dblp.uni-trier.de/db/reference/rsh/rsh2011.html#ShaniG11>