Capstone proposal : Starbucks use-case

Domain background

A company selling products aims to sell even more products, in order to increase its revenue. Various marketing strategies serve this purpose, and one of them is to send offers to customers, in order to encourage them buying more products, or more often.

No matter the strategy, that is always a good thing to understand better the purchasing behavior of your customers. Here, if you have enough customer and transaction data, machine learning can definitely help to discover patterns and structure into big amounts of data.

Problem statement

Starbucks, the well-known coffeehouse chain, implements the offer-sending strategy through an app, on which customers can receive different types of offers. The problem we try to solve is to define which offer would work best for which customer segment; segments will be shaped based on purchasing behaviors and on demographic data.

Datasets and inputs

For this capstone project, 3 data sources are available: data about offers, data about people, and data about transactions. These datasets are nicely provided by Starbucks.

	Filename – entries	**Fields**
offers	portfolio.json – 10	**id** (string) - offer id **offer_type** (string) - type of offer: 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)
demographics	profile.json – 17.000	**age** (int) - age of the customer **became_member_on** (int) - date when customer created an app account **gender** (str) - gender of the customer **id** (str) - customer id **income** (float) - customer's income
transactions	transcript.json – 306.534	**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

In this capstone project, we will try to see if there is a relation between completion rate and customer and offer characteristics. In other words, we will try to see if, based on customer and offer characteristics, we can predict whether an offer will be completed or not by a customer.

If such a relation exists, this is very helpful because Starbucks could leverage our model to send offers only to customers having a high chance to complete the offer.

Benchmark model

Our problem is a classification problem. We will then use the LinearLearner algorithm as a benchmark model. This model is proposed by SageMaker, and able to solve regression and classification tasks.

We will also compare this model with an XGBoost model, also available on Amazon SageMaker. This model reaches very good performance on lots of modern problems, reason why we would like to compare it with the LinerLearner model.

Evaluation metrics

In order to evaluate our model, we will use precision, recall and f-measure. These evaluation measures give a clear and detailed overview of the performance of our model.

Project design

We will complete different steps for this project:

Step 1/ Data exploration: this step will help u familiarize with and understand better the available data.

Step 2/ Data cleaning and data preparation: at this step, we will transform the data in a format usable by our model. Some choices will be made regarding data formatting.

Step 3/ Data modelisation: this step will be the place where we build a first model and evaluate its performance.

Step 4/ Model optimization: based on the performances measured at step 3 and techniques learned during the nanodegree, we will try to improve the performance of our model.

Step 5/ Conclusions: at this step, we will evaluate our results against our initial question. We will also propose possible improvements in order to make our model even more accurate regarding this question.

All these steps will be implemented in a Jupyter Notebook running on Amazon Sagemaker. This notebook will be available on Github afterwards.