

Capstone Proposal

Zhi Deng

Oct 21 2019

Domain background

Starbucks Corporation is an American coffee company and coffeehouse chain. In reality, the Starbucks app sends out various types of promotional offers to customers, either discounts (BOGO or 50% off during happy hours) or Star Dash challenges (completing required purchases to earn star rewards). Sometimes it also informs customers about limited-time drinks, such as those colorful Instagram Frappuccinos. In a simulated environment, Starbucks sends out three types of offers (BOGO, discount and informational) via multiple channels. Customers' responses to offers and transactions are recorded. In either setting, it is important to send the right offer to the right customer.

Problem statement

In this project, I will build a model to predict whether a customer will respond to a promotional offer based on the features of customer and offer.

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

The data is provided by Starbucks and Udacity.

Solution statement

The properties of each offer and customer pair will be combined into a feature vector, while the label on

whether a customer responds to an offer will be the target. Then a binary classification model will be trained to predict the customer's response based on the input feature vector.

Benchmark model

A logistic regression model will serve as the benchmark model in this project. Logistic regression is possibly the most popular algorithm for binary classification problems in industry.

Evaluation metrics

The performance of models will be measured using two metrics, accuracy and F1 score.

Project design

The project will be laid out with the following workflow.

1. Clean data and generate inputs.
 - Process the transcript to collect customers' responses. This step might need careful attention since the transaction pattern in each type of offer may be vastly different, and whether the customer has viewed the offer or not also makes a difference.
 - Join customer and offer features with offer response processed from the transcript.
2. Split data into training set and test set.
3. Perform EDA and feature engineering, build data transformation pipeline if necessary.
4. Train the benchmark model.
5. Train other classification models and select the algorithm with optimal performance.
6. Fine-tune the hyperparameters of selected algorithm.
7. Measure the performance of optimal model and benchmark model using the test set.

References

- [Starbucks - Wikipedia](#)
- [Logistic Regression - scikit-learn](#)
- [Accuracy score - scikit-learn](#)
- [F1 score - scikit-learn](#)