

Domain background

Offer optimization is an interesting marketing problem for any consumer product based company and can have a direct impact on the revenue and the cost incurred by the company. Sending less offers would remove competitive edge and could result in less activity from customers whereas sending too many offers can be overkill and might even have a negative effect on the customer. It is important to provide the right offer to each user at an individual level to maximize profits.

Historically marketing has been done at a universal level across all customers without any personalization. However, this is really inefficient and these days personalized marketing campaigns are becoming more prevalent. Personalized marketing is done by understanding the different groups of customers through data and providing them with the right offer at the right time. One of the ways in which this problem is solved is through customer segmentation. For example, segmenting customers based on demographics and providing a different offer to each of these segments could help drive revenues for a company. Sophisticated segmentation could further be done on the order patterns of the customers using unsupervised techniques such as kmeans to make the offer strategy more effective. However, even segmentation is limited in its strength since the customers within a segment also could be different and might require different offers. To address this issue a more supervised machine learning approach is used where using historical data whether a given offer would be effective on a customer is predicted. This is the approach I would be using as well.

Through this project I hope to gain insights on the patterns of consumers in a consumer product company. I believe as I go through solving this problem, I would understand the salient features of the data and the assumptions that I need to keep in my mind. I would also understand the techniques that work on this kind of data which would give me the confidence to solve future problems in the same domain and help me become a holistic data scientist. I will also be joining as a data scientist in the marketing tech team of a food delivery company early next year, hence I feel this project would be a great learning opportunity for me.

Problem Statement

The aim of this project is to predict which users would respond to a particular offer. Since this is a classification problem the metric that can be used to evaluate the model can be accuracy. Also, since the cost of a false negative (Not giving an offer given that the customer would have responded positively) high here a combination of precision and accuracy also can be considered as a metric.

Datasets and Inputs

The datasets I will be using contain information on 17000 users and their activities with the Starbucks app over a period of 24 days. The activities include transactions made by the user, offers received by the user, offers viewed by the user, and offers completed by the user. An offers dataset provides details on the different offers that have been sent to the users regarding their durability and difficulty. A users dataset provides information about their demographics. The same offer could be offered to a user multiple times during this period. In these cases, I

consider only the last offer that was sent to the user in the case when this user-offer combination hasn't been effective even once. For cases where a user-offer combination is effective I consider the first offer that was sent to the user. An effective offer is an offer that has been viewed and completed by the user. The following are the snippets of the datasets I have

```
In [2]: portfolio.head()
```

```
Out[2]:
```

	channels	difficulty	duration	id	offer_type	reward
0	[email, mobile, social]	10	7	ae264e3637204a6fb9bb56bc8210ddfd	bogo	10
1	[web, email, mobile, social]	10	5	4d5c57ea9a6940dd891ad53e9dbe8da0	bogo	10
2	[web, email, mobile]	0	4	3f207df678b143eea3cee63160fa8bed	informational	0
3	[web, email, mobile]	5	7	9b98b8c7a33c4b65b9aebfe6a799e6d9	bogo	5
4	[web, email]	20	10	0b1e1539f2cc45b7b9fa7c272da2e1d7	discount	5

```
In [3]: profile.head()
```

```
Out[3]:
```

	age	became_member_on	gender	id	income
0	118	20170212	None	68be06ca386d4c31939f3a4f0e3dd783	NaN
1	55	20170715	F	0610b486422d4921ae7d2bf64640c50b	112000.0
2	118	20180712	None	38fe809add3b4fcf9315a9694bb96ff5	NaN
3	75	20170509	F	78afa995795e4d85b5d9ceeca43f5fef	100000.0
4	118	20170804	None	a03223e636434f42ac4c3df47e8bac43	NaN

```
In [67]: transcript.head()
```

```
Out[67]:
```

	event	person	time	value
0	offer received	78afa995795e4d85b5d9ceeca43f5fef	0	{'offer id': '9b98b8c7a33c4b65b9aebfe6a799e6d9'}
1	offer received	a03223e636434f42ac4c3df47e8bac43	0	{'offer id': '0b1e1539f2cc45b7b9fa7c272da2e1d7'}
2	offer received	e2127556f4f64592b11af22de27a7932	0	{'offer id': '2906b810c7d4411798c6938adc9daa5'}
3	offer received	8ec6ce2a7e7949b1bf142def7d0e0586	0	{'offer id': 'fafdc668e3743c1bb461111dcafc2a4'}
4	offer received	68617ca6246f4fbc85e91a2a49552598	0	{'offer id': '4d5c57ea9a6940dd891ad53e9dbe8da0'}

Solution Statement

In order to provide the right offer to the right user, I will be treating this as a classification problem where given the user attributes and offer attributes I will predict whether the offer was effective or not to that particular user. Since this is a classification problem we can use accuracy or f1 score or auc score of the model to determine how well we are doing.

Benchmark Model

This is an unbalanced classification problem as there are more instances of an offer not being effective. I have randomly split the data into train and test; I will be using the latter dataset to evaluate my model and decide the best one. The benchmark model I propose is a naive model where we just predict the majority class for all the data points in the test dataset. This results in an accuracy of 66%. My aim would be to improve upon this naive model with more advanced models and appropriate feature engineering.

Evaluation Metrics

As mentioned previously I would be using a combination of accuracy, precision, f1-score, and auc score to evaluate my models and determine the best one.

Project Design

The following are the steps I have in mind to solve the given problem -

1. Data exploration - In this step I will get comfortable with the data and generally get an idea on the transacting patterns of the users. I will use this exploration to understand the limitations of the data and the assumptions I would need to make to solve the problem. This step would also inform the features that would go into my models.
2. Data Cleaning - In this step I would make any assumptions required and create a data set in which a machine learning model can be used without any issues.
3. Feature Engineering - I will create features which I think will be useful signals that determine whether an offer is effective or not. The features would contain both user and offer related variables. Examples of features I have in mind I have so far are - total amount spent till offer receive time, number of orders till offer receive time, if an offer has been effective on a user in the past etc.
4. Model creation and evaluation - Once I have created the final dataset with the features, I will divide it into train and test. I will train different models on the train data and evaluate it using the test data. Depending on the results I will go back to step 3 to add more important features and then recreate the models.

References

1. <https://towardsdatascience.com/starbucks-offer-optimisation-cdf9bcedd48a>
2. <https://medium.datadriveninvestor.com/offer-optimization-using-machine-learning-46a1f5d1b59b>