Machine Learning Engineer Nanodegree

Capstone Proposal

Atsushi Kawai
June 14th, 2020

1. Definition

1.1 Project Overview

Digital marketing is an efficient way to target specific customers to motivate their behavior. In this project, I use Starbucks rewards mobile app data to analyze and utilize it for the development of digital marketing methods.

Starbucks was started in 1971 in Seattle as a small coffee store, which is affiliated with super market. In 1980's, it launched coffee bar style shop in U.S and develops its style around the world[1]. Nowadays, Starbucks is one of the most well-known coffee chain shop. To serve the products effectively, Starbucks company also use target-marketing with identifying customer segments and using its resource efficiently.

One of Starbucks' approach of marketing is using mobile app data for digital marketing method. Digital marketing is every way to do marketing, if it is regarding with electronic tools or online[2].

Digital marketing can adapt current situation quickly, in which internet usage is growing out among adult generation, by analyzing and profiting from them. Also, we shouldn't forget expansion of internet usage means the past marketing strategy in offline cannot catch up current customer's trend.

1.2 Problem Statement

Starbucks is handling mobile app data for actuating coupon or advertisement to each customers. It has two difficulty to work on. First, schedule and content of offer has limitless possibility. Business developer has so much chance to construct effective offer, as it would be too complicated to decide one specific strategy. Second, it is hard to predict customer's behavior and motivation by seeing through only data acquired from mobile app. To understand these two problems, we need to judge what successful digital marketing is and whether customer behavior is affected by digital marketing or not.

1.3 Metrics

Evaluation metrics I choose in this project is accuracy, precision, recall and F1 score, which all of them are metrics for classification.

- Accuracy : Percentage of correctly classified datasets in total predicted data.

$$Accuracy = \frac{CorrectlyClassifiedPoints}{AllPoints}$$

- Precision : Percentage of guessed positives.

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives}$$

- Recall : Percentage of positives.

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives}$$

- F1 Score : Harmonic mean of precision and recall. This will show whether one of its score is particularly low or not.

$$F1Score = \frac{2 \times Presicion \times Recall}{Precision + Recall}$$

2. Analysis

2.1 Dataset and Data Exploration

The dataset I used in this project is obtained from udacity machine learning engineer course. This datasets is simulated data which mimics customers behavior who use starbucks mobile app.

Advertisement or coupon offer, which is explained in portfolio.json data, has 10 types and these offers are different in the content of offer and way to distribute.

Customer data has two types, profile.json data and transcript.json data. Both data shows customer’s profile and behavior with their schedule.

portfolio.json

- id(string) - offer id
- offer_type(string) - type of offer (BOGO(Buy one, Get one), 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)

	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
5	[web, email, mobile, social]	7	7	2298d6c36e964ae4a3e7e9706d1fb8c2	discount	3
6	[web, email, mobile, social]	10	10	fafdcd668e3743c1bb461111dcafc2a4	discount	2
7	[email, mobile, social]	0	3	5a8bc65990b245e5a138643cd4eb9837	informational	0
8	[web, email, mobile, social]	5	5	f19421c1d4aa40978ebb69ca19b0e20d	bogo	5
9	[web, email, mobile]	10	7	2906b810c7d4411798c6938adc9daaa5	discount	2

Data Speculation

- There are 3 types of offers, BOGO, discount, informational.
- BOGO・・・Difficulty and rewards are always similar. There are 2 types of duration.
- Discount・・・Reward is always smaller than difficulty.
- Informational・・・Reward is zero. Duration is different by each channels.

profile.json

- age(int) - age of the customer
- became_member_on(int) - data when customer created an app account
- gender(str) - gender of the customer
- id(str) - customer id
- income(float) - customer's income

	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

Data Speculation

- There are Nan data.

transcript.json

- event(str) - record description(transaction, offer received, offer viewed, offer completed)
- person(str) - customer id
- time(int) - time in hours since start of test
- value(dict of strings) - either an offer id or transaction amount depending on the record

	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': '2906b810c7d4411798c6938adc9daaa5'}
3	offer received	8ec6ce2a7e7949b1bf142def7d0e0586	0	{'offer id': 'fafdcd668e3743c1bb461111dcafc2a4'}
4	offer received	68617ca6246f4fbc85e91a2a49552598	0	{'offer id': '4d5c57ea9a6940dd891ad53e9dbe8da0'}

Data Speculation
- Similar customers have done transaction several times.
- Some customers respond to the offer they accepted, others not.

2.2 Statics and Exploratory Visualization

After I merged these datasets by customer id and offer id, I have checked each statics and data distributions of customers' action when they received an offer.

Number of data (Number of offer distributed)
- 66628

Age

Mean	54.36
Standard Deviation	17.40
Min	18.00
25%	42.00
50%	55.00
75%	66.00
Max	101.00

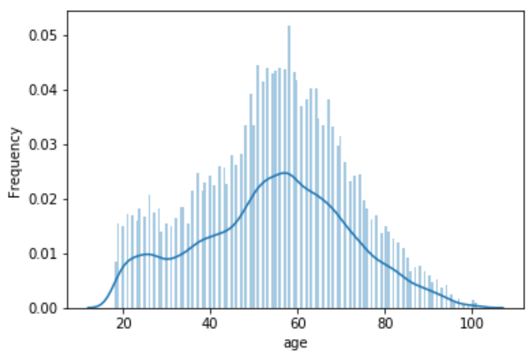

Figure 1. Data Distribution of customer's age

- Data is distributed in wide generation.
- Number of customer's action in 20-30 age are slightly large.

Income

Mean	65329.41
Standard Deviation	21560.41
Min	30000.00
25%	49000.00
50%	64000.00
75%	79000.00
Max	120000.00

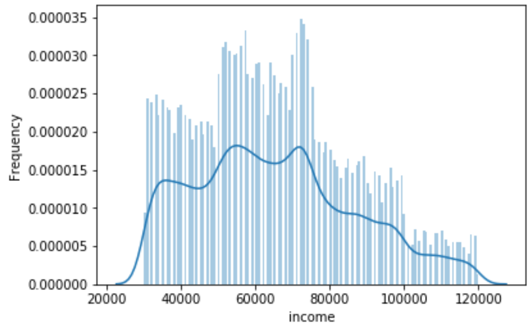

Figure 2. Data Distribution of customer's income

- Income distribution shows several bumps around 30000-80000.
- Distribution looks skewed.

Month (When they received an offer)

Mean	6.68
Standard Deviation	3.49
Min	1.00
25%	4.00
50%	7.00
75%	10.00
Max	12.00

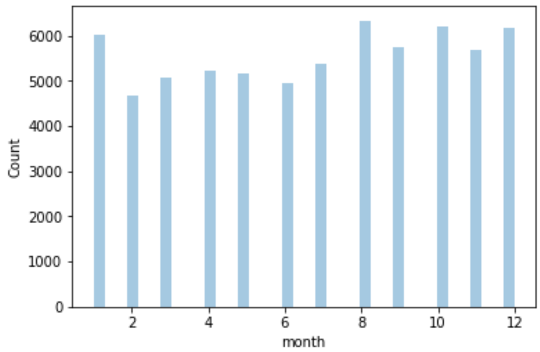

Figure 3. Data Distribution of month

- Number of customers' action is equal in each month.

Year (When they received an offer)

Mean	2016.62
Standard Deviation	1.19
Min	2013.00
25%	2016.00
50%	2017.00
75%	2018.00
Max	2018.00

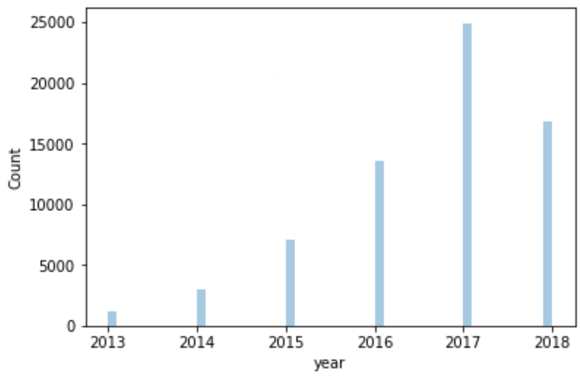

Figure 4. Data Distribution of year

- Number of registration of mobile app seems increasing.

Time (How long time has takes when they received an offer)

Mean	332.44
Standard Deviation	196.52
Min	0.00
25%	168.00
50%	408.00
75%	504.00
Max	576.00

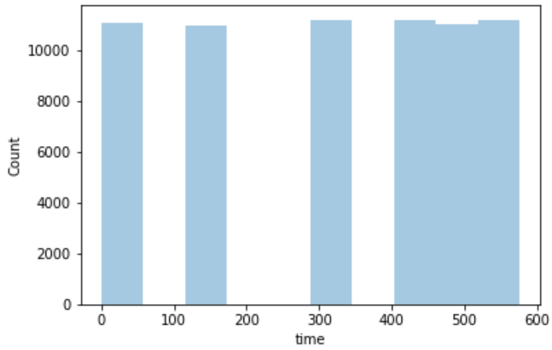

Figure 5. Data Distribution of time

- The distribution of time shows a discrete pattern. It may distribute uniformly .

Duration (How long each BOGO or discount offers have persist)

Mean	180.18
Standard Deviation	58.06
Min	72.00
25%	168.00
50%	168.00
75%	240.00
Max	240.00

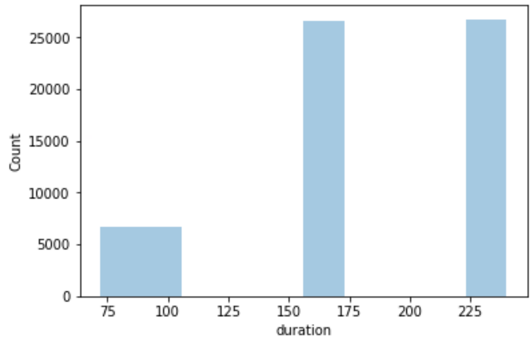

Figure 6. Data Distribution of duration

- Duration of offer shows 4 general pattern.

Reward (How much each BOGO or discount offers would discount)

Mean	2.40
Standard Deviation	1.62
Min	0.00
25%	2.00
50%	2.00
75%	3.00
Max	5.00

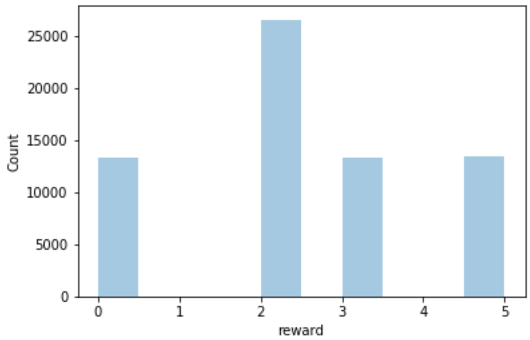

Figure 7. Data Distribution of duration

- Reward of offer shows 4 general pattern.

Pairwise plot of data regarding offer

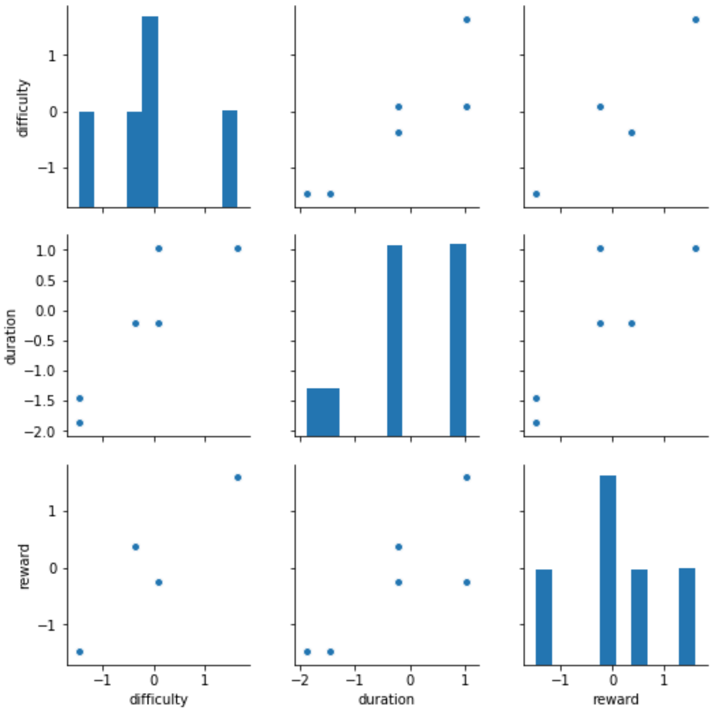

Figure 8. Pairwise plot of data regarding offer

- Data regarding offer properties after standardization shows correlation. But offer property is constructed by company. It is not affected by customer’s features. So, I will include all features for training model.

2.3 Algorithms and Techniques

From data exploration and visualization, there may be unseen underlying pattern in dataset, but basically, most of data distributes uniformly or normally.

The purpose of this project is to figure out which properties of customer or features of offer is important for successful digital marketing and to predict whether each offer can motivate each customers to buy the product.

To do so, at first, I will label successful offer which offer lets customer transact the product, or advertisement is viewed by customer. When I success to label successful offer to each customer in each time schedule, next I will try machine learning to construct the model, which can show feature importances and can predict whether offer is succeed or not. The accuracy of model prediction is evaluated by evaluation metrics, which I will summarize below.

The model I select in this project are XGBoost Classifier and neural network, which I have tried in plagiarism detector project. Both machine learning model can adapt large non-linear dataset.

2.4 Benchmark

To compare and analyze the results of machine learning, I will use naive classifier and logistic regression as benchmark models. Naive Classifier is constructed by setting all label as success. The classification problem could be imbalanced problem, so, by comparing with benchmark models, I can evaluate the accuracy of the results with underlying problem.

3. Methodology

3.1 Data Preprocessing

Preprocessing

- Set datatypes to process.
- Rename feature column name to merge with other datasets.
- Convert duration of offer from date to hours, to compare with mobile app schedule of customer in similar time scale.
- Preprocess date time, when customer become mobile app member, for month and year.
- Drop or fill Nan data.

Standardization

- Standardize customer's age, income, month and year to become member, offer's difficulty, duration and reward.

One-hot-vector processing

- Transform offer distribution channels, offer types, customer's gender as one-hot vector.

Label data for classification

- To label successful offer or not, compare offer duration and the term between the time of action taken by a customer and time of receiving offer, and if it is below offer duration, offer has succeed.

1. If a received offer is completed in a set duration by a customer, label a success as 1 in BOGO dataset.
2. If a received offer is completed in a set duration by a customer, label a success as 1 in discount dataset.

3. If a received offer is viewed in a set duration by a customer, label a success as 1 in informational offer dataset.

Label	Count
0	25019
1	41609

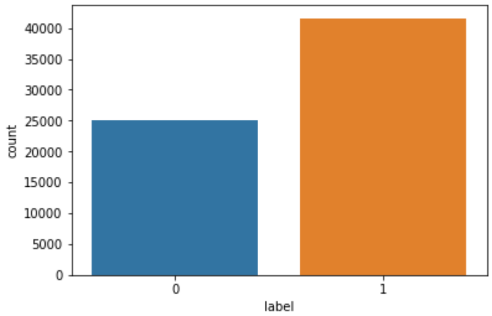

Figure 9. Barplot of label

- Data label is imbalanced, but I will leave it as a data property.

Split the dataset

- To evaluate machine learning model, I split the dataset as training dataset, validation dataset and test dataset in 6 : 2 : 2 proportion.

3.2 Implementation

XGBoost Classifier (Hyperparameters)

- max_depth : 5
- eta : 0.2
- gamma : 4
- min_child_weight : 6
- subsample : 0.8
- silent : 0
- objective : binary:logistic
- early_stopping_rounds : 10
- num_rounds : 500

Pytorch neural network model for classification (Hyperparameters)

- Initial layer - 2 hidden layers - Output layer
- ReLU function as an activation function
- Dropout : 0.2
- input_features : 18
- hidden_dim : 300
- output_dim : 1
- epochs : 10
- batch_size : 1000

3.3 Refinement

To try to improve training model, I used hyper parameter tuning in XGBoost Classifier.

<u>XGBoost Classifier (Hyperparameter tuning)</u>

- max_depth : 3-12
- eta : 0.05-0.5
- gamma : 0-10
- min_child_weight : 2-8
- subsample : 0.5-0.9
- silent : 0
- objective : binary:logistic
- early_stopping_rounds : 10
- num_rounds : 500

4. Results

4.1 Model Evaluation and Validation

To evaluate these models, I have used test datasets to The results of each model is summarized in the table below. From this result, I will select XGBoost Classifier with hyper parameter tuning as the best trained model.

	Accuracy Score	F1 Score	Precision	Recall
Naive Classifier	0.6202	0.7655	0.6202	1.0
Logistic Regression Classifier	0.6791	0.7658	0.6995	0.8460
XGBoost	0.7291	0.7935	0.7526	0.8390
XGBoost (Hyperparameter Tuned)	0.8511	0.8813	0.8718	0.8911
Neural Network	0.7126	0.7791	0.7444	0.8173

4.2 Justification

Even dataset is imbalanced in original dataset, XGBoost Classifier with hyper parameter tuner classified predict successful offer in high accuracy from seeing accuracy score and F1 score.

Also, when we compare precision and recall score between naive classifier and XGBoost classifier with hyperparameter tuner, we can speculate the model doesn't too sensitive to imbalanced data.

5. Conclusion

5.1 Visualization

With this model, we can predict whether an offer successes or not from the character of customer, time after starting mobile app and offer features. Also, we can check feature importance of machine learning model.

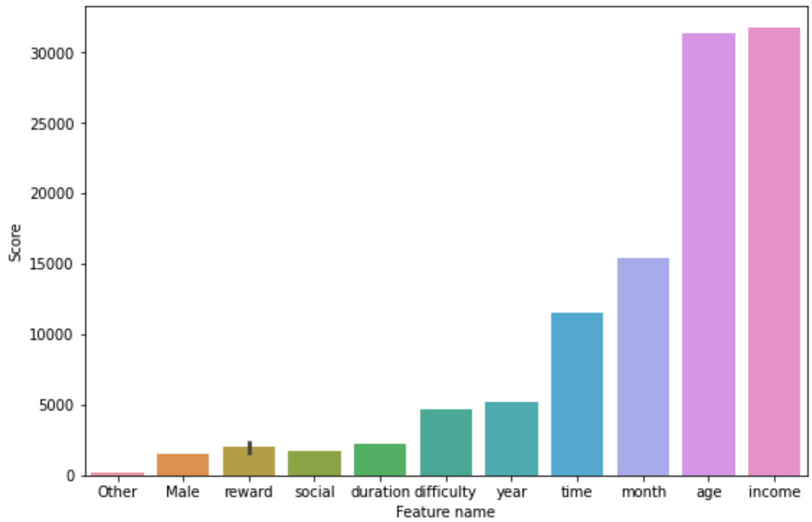

Figure 10. Feature importance of XGBoost model

From this figure, we can see customer features, especially age, income and starting date of mobile app is important for successful offer. Actually, offer property is not so important for successful offer. So, we need to select customers for successful offer.

5.2 Reflection and Improvement

- I have a concern about how to adapt machine learning for imbalanced classification problem. XGBoost model learns nicely, but it is difficult to evaluate from only metrics. Maybe I need to balance imbalanced dataset.
- I cannot success to train neural network well. I think it is because I cannot figure out the best model architecture for it. I would like to know how to construct nice neural network.
- I am interested in the result when I didn't include any offer property features.
- I feel this accuracy and f1 score are well enough. But I am unsure it really it is.

6. Conclusion

[1] Starbucks Japan, https://www.starbucks.co.jp/company/history/fy2000.html
[2] Lucy Alexander, "The Who, What, Why, & How of Digital Marketing", https://blog.hubspot.com/marketing/what-is-digital-marketing, 2020.