Starbucks Capstone Challenge

Project Proposal

Domain Background | Problem Statement | Datasets and Inputs | Solution Statement

Benchmark model | Evaluation metrics | Project Design

Sooyeon Won

21. January 2021

won.sooyeon@live.com

Domain Background

This project is derived from the field of Customer Relationship Management (CRM) in Starbucks. One of the main concerns for CRM field is to interact with not only current customers, but also previous and potential customers, so that a company can improve its business relationship with all customers, and eventually expect its continuous sales growth. Nowadays, there are various channels that a company can manage its customers such as using mobile application, company's website, emails, and so on. [1] Recently I am employed as a CRM data analyst at a retail industry. This could be one reason, I chose this project, which improves my future professional skills further.

Starbucks Corporation is a multinational coffee company, whose headquarter is in Seattle. It serves all kinds of coffee-related products.[2] Starbucks is often evaluated as the leader on the highly substitutable market. Its customers realize that they not solely purchase a coffee at Starbucks, but also enjoy the unique atmosphere at Starbucks store. Moreover, Starbucks is well-known for its own solid online marketing program to manage its global customers. [3]

As mentioned in the provided business case, one of its marketing campaigns is to send out an offer to customers through various channels. An offer can be either just an advertisement for a certain beverage (**informational offer**), or a coupon-type offer such as a '**Discount**' or 'buy 1, get 1 (**BOGO**)'.

- **Informational offer**
  Its main purpose is providing an (updated) information about products to customers. There is no reward, neither a required spending that a customer is expected to spend.

- **Discount offer**
  Customers get monetary rewards which is equivalent to a certain proportion of the amount of spending.

- **Buy 1, Get 1 (BOGO)**
  Customers should spend certain amounts (threshold) so that they can receive monetary rewards, which is equal amount of the cut-off amount.

Each offer is valid for certain number of days. The validity periods are different from offers.

Problem Statement

Not all of customers obtain the offers, and different customers receive different the type of the offers. This can be differ according to the demographic factors or individual purchasing patterns. As a CRM analyst, we should maximize a return-on-investment (ROI), since all marketing campaigns have related-costs. To do so, I identify the desirably-used offers in this analysis. Then I analyze the customers who used the offers desirably, which features do the customers share each other.

To be specific, if a customer segment which is likely to react regular offers, so that the offers make the customers regularly purchase drinks at our chains, then our marketing events should focus on the customers, rather than customers who show frequent purchasing patterns regardless of product offers.

Additionally, if a customer feels reluctant to receive advertisement offers, this might be able to result in losing users or lead to decrease customer retention rate in the long run. These types of customers should be removed in advance from the “offering-list”.

In this analysis, I classify the customers into 2 groups who use appropriately our offers and who do not, based on customers' individual demographic features and their purchasing patterns.

Datasets and Inputs

To solve the business issues, Starbucks provides 3 datasets:

Profile

- Rewards program users
- Format: json
- Size: 17,000 users x 5 features
- Included Features:
    - gender: (categorical) M, F, O, or null
    - age: (numeric) missing value encoded as 118
    - id: (string/hash)
    - became_member_on: (date) format YYYYMMDD
    - income: (numeric)

Portfolio

- Offers sent during 30-day test period
- Format: json
- Size: 10 offers x 6 features
- Included Features:
    - reward: (numeric) money awarded for the amount spent
    - channels: (list) web, email, mobile, social
    - difficulty: (numeric) money required to be spent to receive reward
    - duration: (numeric) time for offer to be open, in days
    - offer_type: (string) bogo, discount, informational
    - id: (string/hash)

Transcript

- Event log, basically transaction records
- Format: json
- Size: 306,648 events (transactions) x 4 features
- Included Features:
    - person: (string/hash)
    - event: (string) offer received, offer viewed, transaction, offer completed

- value: (dictionary) different values depending on event type
  - offer id: (string/hash) not associated with any "transaction"
  - amount: (numeric) money spent in "transaction"
  - reward: (numeric) money gained from "offer completed"
- time: (numeric) hours after start of test (experiment)

Solution Statement

Note that it is appropriate when each stage in the ‘event’ field of “Transcript” dataset occurs in consecutive order.

A Possible Process of Using Offers

Case	Received	Viewed	Transaction	Completed	Reward	Desirable
1	√	√	√	√	√	Yes
2	√	√	√			Yes
3	√	√				No
4	√		√	√		No
5	√		√			No
6	√					No
7			√			No

However, the prior stages are not necessarily required to complete the offers. For example, a customer who did not view the offer, but the customer could complete to use the offer then get the corresponding rewards, though. Based on this fact, I will filter such customers who are not influenced by offers, because their purchases are independent from the fact whether they received the offers. (Case 4, Case 5, Case 6, Case 7)

I also define the Case 3 as ‘undesirably-used’ offer cases, because the customers viewed the offers but this does not drive to further transaction stage. Therefore, Case 1 or Case 2 is defined as desirably-used offer cases.

After then, I will conduct the analysis using frequently used Classification models with Random Forest- , AdaBoost-, and Gradient Boosting Classifier. Additionally, XG Boost -, LightGBM and CatBoost Classification models will be applied as complimentary supervised learning approaches.

Then I will compare the performances of each model. Finally, I will select one which showed the best performance. The target variable in this analysis is the way that each customer used offers desirably. In addition, customer demographic features and RFM purchasing characteristics will be used as input variables. This classification analysis is conducted for each type of offer separately.

Benchmark model

Although the models which show less performance than the model with best performance could be considered as benchmark models, I designated a "logistic regression" as a benchmark model for this project to compare my solution objectively.

Evaluation metrics

The performance of individual models can be measured according to various evaluation metrics. As mentioned earlier, I will solve this business case by applying machine learning classification models which is a type of supervised machine learning techniques. This is because we are trying to figure out whether each offer is effectively used by customers, based on their demographic features and purchasing patterns.

I will firstly compute Accuracy, Precision, Recall, and F1-Score. "Accuracy" however is not appropriate to be used as an evaluation metric for predictive models when classifying in predictive analytics. As "Accuracy Paradox" indicates, a simple model might be able to achieve a high score of accuracy but be too crude to be useful. [4] On the other hand, F1-Score is computed with the prediction and recall of the test.

$$F1\ Score = 2 * \frac{Prediction * Recall}{Prediction + Recall}$$

Since the F1-Score indicates a weighted average of the prediction and recall values, it takes false positive as well as false negatives into account. F1-Score reaches its maximum (best) value at 1 and its minimum (worst) value at 0. F1-Score is often useful in comparison to accuracy, especially if each class is not evenly distributed. Thus, I will select the best performance model according to its F1-Score. [5]

Evaluation Metrics	Meaning
Accuracy	The proportion of correctly predicted cases.
Precision	The number of true positive cases over the number of true positive - and false **positive** cases.
Recall	The number of true positive cases over the number of true positive - and false **negative** cases.
F1 - Score	It is the harmonic mean of precision and recall. Higher scores indicate better performance.

Project Design

My approach to the solution is as follows.

1. Introduction

- Brief explanation of current business case
- Clarifying the final goal of this project

2. Data Preparation/ Cleaning

Conducting the below usual data wrangling processes for the given datasets: Portfolio Dataframe, Profile Dataframe, Transcript Dataframe

- Deleting duplicates, Imputing missing values
- Missing values in the 'age' column will be replaced with its median value
- Outliers: Check whether the extreme values are realistic. If unrealistic, outliers will be removed by taking the datapoints from 5th percentile to 95th percentile[1]
- Adjust the datatype of each feature, if needed
- Tidiness of dataset: One dataset should contain one information.

3. Data Exploration

- Exploring the current business situations
  - The Change of traffics during the test period
  - The Sales Trend Amount during the test period
- Exploring demographic features of current customers
  - Any statistically significant relationships between issued offer-types and gender, age, income, number of days as a Starbucks member
- Exploring Starbucks Offers
  - How the offers are distributed during the test period
- Exploring Customer Purchasing Patterns
  - RFM Analysis

4. Data Analysis

- Feature Engineering
  - Define the 'desirably-used' offers (Target variable)
  - Assign each transaction to 2 classes: 'desirable', 'non-desirable'
  - Imbalanced Data, Feature Selection
- Data Splitting: Training, Validating, Testing Data
- Data Balancing: Synthetic Minority Oversampling Technique (SMOTE)
- Model Comparison: sklearn ensemble methods, and more
- Model Selection
  - Evaluation Metrics: Accuracy, Precision, Recall, F1-score
  - Model Selection Metrics: f1-score

5. Conclusion

- Compare several input features that show significant difference between 'desirably' predicted samples and 'non-desirably' predicted samples.

[1] The percentile can be changed, depending on the size of dataset.

References

[1] https://en.wikipedia.org/wiki/Customer_relationship_management

[2] https://de.wikipedia.org/wiki/Starbucks

[3] https://www.ukessays.com/essays/marketing/starbucks-the-company-philosophy-marketing-essay.php

[4] https://towardsdatascience.com/accuracy-paradox-897a69e2dd9b

[5] https://blog.exsilio.com/all/accuracy-precision-recall-f1-score-interpretation-of-performance-measures/