Machine Learning Engineer Nanodegree - Udacity

Capstone Proposal

Konstantinos Trikaliotis - January 2022

Domain Background

This project stems from the marketing system used by Starbucks to stay in touch with its customers. Starbucks is America's largest coffee shop chain and a major organization that wants to attract new customers and motivating existing customers to consume more products by offering special offers and promotions through personalized targeted offers. The main goal is to incentivize and reward customers who are registered on its platform by periodically sending individual messages / notifications through multiple channels such as email, social networks, web or directly through the Starbucks app. These messages / notifications contain offers related to its products and are divided into three main categories:

- BOGO: Buy-one-get-one is a type of voucher where users can get a reward equivalent to a threshold amount if they spend a certain amount of money.
- Discount: In the discount offer, the user earns a reward equal to a fraction of the amount spent.
- Information: In an informational offer, there is no actual reward, but also no mandatory amount the user is needed to spend.

Therefore, companies want to increase the profits generated by these marketing campaigns by implanting recommendation platforms/algorithms which will suggest what kind of vouchers should be offered to each individual customer, by looking at several factors which would lead to a succefull offer. A simple definition of a successful offer is made up of two characteristics:

1. The customer has received, viewed, and completed the offer
2. The Customer’s spending will increase after or during the offer period.

An example of a failure would be a customer who would buy a product of $10 without seeing any offer and would receive a 20% discount on that amount due to an unseen offer received. This is not a successful offer at all, as the customer receives the offer without changing buying behaviour or increasing spending, and as a result the company loses profits.

Considering all the details that must be observed by a company's marketing team, we can understand the difficulty of this matter. As a result, advanced machine learning recommendation algorithms are an essential tool for any marketing team, as they are able to extract historical insights from customers’ behaviour and determine the most relevant offer for each one of them, by just looking at some basic transaction and demographic data.

I personally consider this a really fascinating area, as providing better marketing campaigns can not only benefit businesses by increasing their profits, but also the customers who would receive more relevant incentives based on their consumption behaviour. Perhaps this win-win is the key to sustaining economic growth while people can afford better services and products.

I have found some interesting articles, blogs and repositories that have helped me a lot in making suggestions for this project. These are in the reference section below.

Problem Statement

Any huge enterprise such as Starbucks, invests an excessive amount of cash in advertising campaigns expecting to increase its profits through attracting new clients or by having a higher profit on the same existing customers than the assumed earlier. Therefore, it is important to identify the most relevant offers to the right customers for a successful campaign.

Nevertheless, a category of customers may not even see the offer provided to them. This is really likely to be a problem with the selected channel but may be related to other demographic characteristics of the customer (e.g. old customers do not want to try new products even if there are offers related to them).

Moreover, there is another category of customers who buy nothing, despite seeing an offer. This might be due to a problem with the offer sent, or it may not be a customer that should be considered as a target for this specific voucher type.

In other cases, customers may empathize with the offer, try new products, or spend more than usual and these are the situations that need to be identified, tracked and suggested by a model, platform or algorithm in general.

The problem this project is trying to solve is to identify / predict the most appropriate offer for each individual customer. That is, finding an offer that is likely to get your customers to buy more Starbucks products and increase the organization's profits.

In the context of this project the definitions of an appropriate and not appropriate offer are:

- An appropriate offer is that one where the consumer sees the offer provided and buys merchandise under its influence, finishing the offer lifecycle.
  Offer Received -> Offer Viewed -> Transaction occurred -> Offer Completed

- Not appropriate offer examples:
  - Customer does not see an offer
  - Customer sees the offer but does not complete it, as the offer did not force or convinced the customer to buy the product
  - Customer purchases some goods and completes an offer and receives a reward before viewing that offer, as the consumer was not affected by this offer when he / she decided to make a transaction.

Datasets and Inputs

The dataset used in this project is provided by Udacity and Starbucks as part of the Machine Learning Engineer Nanodegree program. Starbucks provides simulated data that mimics customer behaviour on their rewards mobile app.

More specifically, the data is a simulation of how people make purchasing decisions and how those decisions are driven by advertising offers. The program used to create the data simulates how people make purchasing decisions and how those decisions are influenced by advertising offers. Not all customers get the same offer, and this is the challenge will try to solve with this data set.

Every purchase decision completed by each customer is influenced by some hidden characteristics. Consumers produce numerous events, including receiving offers, opening offers, and making purchases.

For simplicity, there are no explicit products to track, but only the amount of each transaction or offer is recorded.

The data is contained in three files:

- **portfolio.json** - *containing offer ids and meta data about each offer (duration, type, etc.)*
    - id (string) - offer id
    - offer_type (string) - type of offer i.e. BOGO, discount, informational
    - difficulty (int) - minimum required spend to complete an offer
    - offer_type (string) - type of offer ie BOGO, discount, informational
    - reward (int) - reward given for completing an offer
    - duration (int) - time for offer to be open, in days
    - channels (list of strings)

- **profile.json** - *demographic data for each customer*
    - age (int) - age of the customer
    - became_member_on (int) - date when customer created an app account
    - gender (str) - gender of the customer (note some entries contain 'O' for other rather than M or F)
    - id (str) - customer id
    - income (float) - customer's income

- **transcript.json** - *records for transactions, offers received, offers viewed, and offers completed*
    - event (str) - record description (i.e. 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

To address the above problem, this project suggests implementing a machine learning model able to analyse and extract useful information about the behaviour of the customer, by analysing the transcription data of their interaction with the Starbucks app.

The problem will be approached, as a binary classification problem, where the label is 1 if a customer is eligible for a specific offer and 0 if he / she is not. Moreover, the project will investigate if it is possible to identify and use the information of any money lost from some offers provided. There are specific demographic groups of customers, which have received a discount, without even opening the corresponding offer which means that the customer was not aware of that offer and his / her behaviour did not change (the customer did not increase his / her spending).

The most common algorithms used for binary classification models are SVM (support vector machine), decision tree, random forests, ensemble methods and neural networks. This project will focus on creating an ensemble boosting tree algorithm and more specifically, an XGBoost model which is considered as really powerful and easy to be trained model.

The final goal is to be able to create a marketing system that will help the marketing team at Starbucks to identify the most appropriate offer for each customer.

Benchmark Model

A decision tree classifier will be used as a benchmark model. Both models (decision tree and XGBoost) will be trained and tested at identical datasets. In addition, the same evaluation metrics will be used, so that the results can be compared.

An XGBoost algorithm is an ensemble method, which consists of individual simple models and, more specifically, decision trees. This is why a simple decision tree, which is considered as a really great algorithm for binary classification problems, has been selected as the best benchmark model in our problem.

One potential metric that can be used to measure the performance of both models is the accuracy on both training and test sets. Moreover, it would be good to look at the confusion matrices of both models and identify any huge differences in the false negatives and false positives. The mathematical equation of the accuracy and a confusion matrix sample can be found below.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

		Predicted	
		Negative	**Positive**
Actual	**Negative**	True Negative	False Positive
	Positive	False Negative	True Positive

Evaluation Metrics

As mentioned above, the metric that will be used to evaluate the performance of all models is the accuracy. Moreover, the confusion matrix will be plotted on each model, to identify any potential issues in the false positives or false negatives. By using the same evaluation metrics on all models, we make sure that the comparison between them is quantified and we are comparing “apples with apples”.

The accuracy is one of the most known and most used evaluation metrics and it has been used by many other people completing similar projects. This was one of the main reasons that has been selected as the main evaluation metric on this project too.

Definition of accuracy:

- *Accuracy is one metric for evaluating classification models. Informally, accuracy is the fraction of predictions our model got right. Formally, accuracy has the following definition:*

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$$

- *For binary classification, accuracy can also be calculated in terms of positives and negatives as follows:*

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Project Design

1. **Data Loading and Exploratory Data Analysis (EDA)**
   The data will be extracted and loaded in the correct format in a Jupyter notebook. Then, statistical data visualization will be performed to summarize the main characteristics of each feature / column and identify any statistical issues.

2. **Data Pre-processing**
   i. Data Cleaning: Look at the visualization created on the previous step and identify patterns and any potential issues. Find the best way to address any data issue such as null values on some columns or categorical features that need to be encoded.
   ii. Data Transformation: Find the best way to combine all three tables into one summarized table which will be used as the main data source for features engineering and creating the final training and testing sets.

3. **Feature Engineering**
   Looking at the exploratory data analysis performed in the first step and determining which features are going to be used for predicting the outcome of our dependent variable. In addition, identifying features that can be engineered using the existing independent variables and performing feature reduction methods (PCA, etc .) if needed.

4. **Preparing and splitting the data into training, validation, and testing sets**
   After completing the feature engineering and having the final table which will be used to feed the models, the last step is to split the data into training, validation, and testing sets:
   - *training set:* This is the largest sample dataset and will be used to initially fit all models
   - *validation set:* This sample dataset will be used as an unbiased evaluation set of all models during the training process and will help to define the best hyperparameters for the final XGBoost algorithm
   - *testing set:* This sample dataset will be used for the final comparison of all the models and for the general evaluation of their performance on a real-world scenario. This will be an unseen dataset and will be used once the model is completely trained.

5. **Training and testing the performance of the benchmark model**
   The datasets created in step 4 will be used to train and test the performance of the benchmark model

6. **Constructing a powerful XGBoost model**
   i. Training and testing a random XGBoost classifier, by using some initial hyperparameters
   ii. Hyperparameter tunning, using the training and validation sets during the training
   iii. The final tuned and trained model will be used for evaluating the performance of the model in the testing set

7. **Conclusion**
   i. Comparing and evaluating the results of the final XGBoost and benchmark model by using the accuracy as a main metric and by looking at the confusion matrices for any significant differences
   ii. Presenting final predictions and model usability in a real-world scenario

References

1. Nicolas Guan, "Who Might Respond to Starbucks' offer?"
2. Silvio Mori Neto, "Starbuck's Capstone Challenge"
3. Cheb Wabgm Cgebgyuan Deng, Suzhen Wang, "Imbalance-XGBoost: leveraging weighted and focal losses for binary label-imbalanced classification with XGBoost"
4. Classification: Accuracy
5. Josh Xin Jie Lee, "Implementing a Profitable Promotional Strategy for Starbucks with Machine Learning"
6. Didrik Nielsen, "Why Does XGBoost Win "Every" Machine Learning Competition?"