Machine Learning Nano Degree Capstone Proposal – Starbucks - Ishan Pathak

1. Introduction

Starbucks is a multinational coffee chain which operates in a highly substitutional market. In this problem Starbucks is looking for improving the effectiveness of the interaction which it has with its existing customers. The problem asks for identifying is a customer will successfully respond to an offer rolled out to him/her.

In the current context, Starbucks has considered three diverse types of offers to its customers:

- **Discount Offer** – Discount on next purchase above a certain amount
- **BOGO** – Buy One Get One Free for purchase worth a certain amount
- **Informational** – This has information on existing Starbucks Products

Not all the existing customers react to these offers similarly. Take for example, a regular Starbucks user, in all probability, the offers delivered to this customer would not have a significant impact on the purchase activity for this customer. On the other hand, a customer who is infrequent to Starbucks or using a substitute product in market (say Costa Coffee) could react favorably to some of these offers.

This project is derived from the field of Customer Relationship Management (CRM) in Starbucks. One main concern for CRM is to interact with not only current customers, but also earlier and potential customers, so that the company can effectively keep its business relationship with all customers.

2. Problem Statement

The Problem Statement for this Capstone Challenge is to analyze the dataset for Starbucks Customers and build a model that can predict whether a customer will respond to an offer. For Starbucks, this is an important question since there is a direct marketing cost associated sending out offers to customers. Hence, we would like to roll-out offers to customers where there is a high likelihood for the offer to be desirable. Also, if we are unable to decipher certain good offers to customers then that would represent a loss in acquisition of new customers or the retention of existing customers thereby leading to reduction in expected Starbucks profits.

In the Capstone Project, therefore, our main goal would be to correctly label the desirable and non-desirable offers with which we will train our ML Model. The predictions by the trained model on a test dataset of customers and offers will then help ascertain if a specific type of offer is desirable for the customer or not. We will explain the solution statement in detail in later sections.

3. Datasets and Inputs

To solve the problem at hand, Starbucks provides us with the following three datasets

Profile Dataset

- Details on the Customers participating in Starbucks Programs (receiving offers/viewing offers, conducting transactions, etc.)
- Shape: 17000 x 5
- Features
    - Gender: (string) M, F, Not Available
    - Age: (numeric) missing values encoded as 118
    - Id: (string) customer identification code
    - became_member_on: (date) date on which customer enrolled in Starbucks program
    - income: (numeric)

Portfolio Dataset

- Details on the offers sent during the test period
- Shape: 10 x 6
- Features
    - reward: (numeric) money awarded for the spent amount
    - channels: (list) web, social, email, mobile
    - difficulty: (numeric) minimum money spent to complete the offer
    - duration: (numeric) time for which an offer is valid
    - offer_type: (string) bogo, discount, informational
    - id: (string) offer identification code

Transcript Dataset

- Details on the event logs like offer viewings, offer received, offer completed, transactions conducted
- Shape: 306,648 x 4
- Time: time elapsed in hours after the start of experiment
- Features
    - person: (string) customer id for identifying customers
    - event: (string) offer received, offer viewed, offer completed, transaction
    - value: (dictionary) values depending on the event type
        - offer_id not associated with transaction
        - transaction amount for transaction completed
        - rewards for completed offers

4. Solution Statement

Methodology

In my current implementation, I have first classified all the offers rolled out to customers into one of the two categories:

- **Desirable**: An offer will be classified as desirable in any of the two situations:
  - If the customer receives the offer, then views the offer then conducts a transaction sufficient to complete the offer all within the offer validity period then the offer is desirable
  - If the customer receives the offer, then views the offer then conducts a transaction within the validity period which is insufficient to meet the offer difficulty can still be considered a desirable offer since we can conclude that the transaction was performed under the influence of the offer
- In all other cases, the offer will be classified as **undesirable**

In the implementation, we are even including the offers that could not be completed due to not meeting the minimum transaction thresholds (I.e., offer difficulty) but still after viewing these offers, the customer did buy certain products which can help us conclude that the purchase was done under the influence of the offer. In this case, rolling the offer is leading to an increase in Starbucks sales thereby making it favorable. The task for us at hand then becomes classification of offers into of the above two categories in the data dataset. Note that the analysis is done separately for different classes.

Benchmark Model

We use a benchmark model which is Logistic Regression to perform the classification task. Logistic Regression is a supervised learning algorithm which is used for categorical target variables. Logistic Regression is implemented through sklearn library in Python. In the next step, we implement the SageMaker XGBoost and Linear Learner Model and try to assess the relative performance for different offer types. Finally, depending on the business context, we chose the best model to use.

Evaluation

Performance of the models is assessed through the following evaluation metrics:

- Accuracy
- Precision
- Recall
- F1 Score

For our classification task we need to correctly classify the positives (since errors here would lead to an economic sales loss) as well as the negatives (since errors here would lead to excessive direct marketing costs). Accuracy is the total fraction of correct predictions. Precision tells us the extent to which the positive class predictions have been correct, and recall measures the percentage of positive ground-truths correctly labelled as positive. F1 score is the harmonic mean of precision and recall.

Model selection for prediction is dependent on the business questions which we are trying to answer. False negatives predicted by the model would mean a future economic loss since the customer would have desirably responded to the offer by conducting a transaction. False positives predicted

by the model would mean wasted marketing cost incurred in rolling the offer to the customer. Ideally, we would like to reduce the proportion of both false positives and false negatives.

For business situations which require us to predict the negatives correctly, we use the accuracy score as the deciding criteria whereas for business situations which require us to predict the positives correctly, we use the recall rate as the deciding criteria. We will also use the confusion matrices to infer performance of models under different business contexts.

5. Project Design

1. Introduction

- Explanation of the problem statement, datasets and methodology

2. Data Preparation and Cleaning

- Deleting Duplicated, imputing missing values with appropriate metrics
- Separating the Transcript datasets into two separate datasets involving offer related and transaction related events

3. Data Exploration

- Exploring the business activities during the test period
    - Exploration of trends in average and aggregate consumer activity over test period
    - Exploration of patterns in consumer demographics over sample
    - Exploration of trends in consumer average and aggregate sales amount over the test period
    - Exploration of distribution of offers across the population demographics like age, income, etc.

4. Feature Engineering

- Insertion of "**desirable_use**" feature which classifies the offers rolled in to desirable/non-desirable categories
- Insertion of metrics summarizing the customer purchasing patterns namely—this is based on RFM Analysis
    - **Recency** – How recently has the customer purchased Starbucks products
    - **Frequency** – How frequently does the customer purchased Starbucks products
    - **Monetary** – How much does the customer spend on Starbucks purchases

4. Modelling and generating predictions

- Data splitting into Training, Validation and Test Datasets
- Initial Run of Benchmark Logistic Regression Model
- Initial Run of SageMaker XGBoost Model
- Adjustment of Model Inputs if required e.g., input data balancing
- Rerun of benchmark Logistic Model, SageMaker XGBoost Model and SageMaker Linear Learner Model
- Analysis of Results and choosing the best model for the use-case

6. References

https://en.wikipedia.org/wiki/Gradient_descent

https://en.wikipedia.org/wiki/Logistic_regression

https://docs.aws.amazon.com/sagemaker/latest/dg/xgboost.html

https://docs.aws.amazon.com/sagemaker/latest/dg/linear-learner.html

https://en.wikipedia.org/wiki/RFM_(market_research)