Domain Background

I am choosing project from Udacity recommended project – Starbucks challenge. Project is related to using Data and Machine Learning for manage online **Target Advertising**.

Targeted advertising is a form of advertising that is directed towards an audience with certain traits, based on the product or person the advertiser is promoting.

Targeted advertising is an effective tool for retailers to reach the right audience at the right time, in the right place, and relevant targeting is an important component in any retailer's digital marketing strategy.

Any company that interact with customers through online medium, like selling projects online (E.g. Amazon) or using app for business transaction (E.g. e-wallets) wants to understand their customers, their buying preferences and then advertise products and offers which they actually need at the right time which ultimately makes customer experience more personalized and boost revenues for the marketer.

Examples where ML is being utilized in identifying target audience.

- Effects of Training Datasets on Both the Extreme Learning Machine and Support Vector Machine for Target Audience Identification on Twitter (https://link.springer.com/chapter/10.1007/978-3-319-14063-6_35) – This research shows using two machine learning approaches - the Extreme Learning Machine (ELM) and Support Vector Machine (SVM) in identifying a target audience in order to maximize marketing efficiency and improve the return of investment (ROI).
- Predicting Consumer Behaviour in Digital Market: A Machine Learning Approach (https://www.researchgate.net/publication/335149938_Predicting_Consumer_Behaviour_in_Digital_Market_A_Machine_Learning_Approach) – This paper shows how companies are using data and ML in Digital marketing techniques such as search engine optimization (SEO), search engine marketing (SEM), influencer marketing, content automation, campaign marketing.

Why I have chosen this project?

This is one of example of personalization for a customer which is of great interest to me. As per Forbes (https://www.forbes.com/sites/blakemorgan/2020/02/18/50-stats-showing-the-power-of-personalization/#5610869b2a94), from Spotify to Netflix and Amazon, we're surrounded by extreme personalization every day. Consumers have come to expect that same level of personalization from companies of all sizes. Investing in personalization efforts to build relationships and create better experiences can pay off with serious rewards for brands.

Problem Statement

As Starbucks is concerned with enhances customer experience, Starbucks need to effectively manage marketing campaign it offers to its customer through its app.

When Starbucks advertise promotional offers to its customer, there are different ways in which customer can react.

- Ad can motivate the customer to make a purchase. This can be a positive experience for customer as customer is getting personalized experience.
- If offer is of no interest to customer, he/she is likely to ignore the ad. This is neutral experience for customer as ad has no impact on customer.
- If customer is advertised ads too many times where he/she is not interested, it can be frustrating experience and likely to demotivate customer to make a purchase. This is negative experience for customer and impact Starbucks in negative way.

Starbucks would want to send offer to its customer where a customer is possibly going to react in a positive way and would want to avoid sending any offer to customers who are likely to react in neutral or negative way. Starbucks would also like to understand which offer is best for customer in case customer is likely to

Datasets and Inputs

Here I am using dataset provided by Udacity.

There are 3 data files required for this project.

- portfolio.json - containing offer ids and metadata about each offer (duration, type, etc.)
    - There are 10 offers.
    - 2 are of type informational, 4 are of type BOGO and 4 are of type discount
- profile.json - demographic data for each customer
    - There are total of 17000 customer covered
    - 50% customers are male, 36% customers are female. Data seems to be missing for 13% customer
- transcript.json - records for transactions, offers received, offers viewed, and offers completed
    - There are 4 type of records in this table – ‘offer received’, ‘transaction’, ‘offer viewed’, ‘offer completed’.
    - There is a time variable which represent time at which record was generated.
    - Each offer is send to approx. 6000 customers which shows uniform distribution.

- Distribution of persons for offer type shows that BOGO and Discount offer have ~14000 unique persons and information has ~10000 unique persons which is similar to distribution of offer type in portfolio dataset.

Solution Statement

Here a machine learning model can be used predict whether any customer should be sent any offer or not and if yes, which offer should be send to that particular customer.

Based on customer transactions, numbers of features can be created that reflects the customer spending pattern and demographics and based on that, prediction can be made if customer should be sent which promotional offer.

Output of model can be like this

- 0 – Do not send any offer
- 1 – Send buy-one-get-one (BOGO)
- 2 – Send discount
- 3 – Sent informational

Benchmark Model

As this is a classification problem, there are multiple models that can be utilized here.

For benchmark, we can use a simple linear/logistics model.

Evaluation Metrics

As per investigation on data, it seems that dataset is well balanced for each offer and offer type as well.

Accuracy is easily suited metrics for easily suited for binary as well as a multiclass classification problem.

Accuracy is a valid choice of evaluation for classification problems which are well balanced and not skewed or No class imbalance.

Therefore Starbucks can choose **accuracy** for evaluation of model.

Project Design

Project can be designed in following way.

Using transactions and profile data, features can be created which represent customer demographics, buying patterns and preferences. This can be linked with portfolio data to see how customers buying pattern and preferences changes before and after customer is send offer.

Basis on these features, model can predict whether to send any offer to that customer or not or if yes, which offer is best suited for that customer.

For evaluation, we can divide dataset into train and test dataset. Splitting of dataset can be done uniformly over customer demographics and then we can use evaluation metrics to select best model algorithm.

Since model is going to be trained on 30-day test period for 3 offers, model can be deployed any time Starbucks choose to send these 3 offers on wider customer population.

It is assumed that for test period, appropriate customer population has been selected which is diverse and represent wider variety of customers therefore it would be safe to deploy model over wider customer population.

In case, new offers are designed by Starbucks, then model would be required to be retrained using data of new test period.

In case model does not give satisfactory results, we can still do data exploration analysis to come up with useful insights which can help Starbucks plan and manage their marketing campaigns and offer designs.