Acquire Valued Shoppers

Introduction:

One of the challenges that businesses face today, is to attract and retain customers. Along with providing quality services, making the customers repeatedly buy products/subscribe to services is crucial for success of new businesses. Businesses run huge sales with discount coupons/offers to attract customers.

Problem:

As part of this project I would like to predict if the customers who received the discount offer on a product, will buy same product again.

Problems like these could be solved for companies in B2C/Subscription model to see effectiveness of discount offers on products and identify which company products customers prefer to buy when discounts are offered, what time of the year the offers are effective.

I will be using customer, offer data for over 300000 customers and over 3 million transactions by those customers, this data is available on Kaggle.com

https://www.kaggle.com/c/acquire-valued-shoppers-challenge/data

Files:

- transactions.csv - contains transaction history for all customers for a period of at least 1 year prior to their offered incentive
- trainHistory.csv - contains the incentive offered to each customer and information about the behavioral response to the offer
- testHistory.csv - contains the incentive offered to each customer but does not include their response (you are predicting the repeater column for each id in this file)
- offers.csv - contains information about the offers

Approach:

Step1 - I will look at customers who received the offer(history) and offers information data without looking at the transaction data, this will give a base line to start.

I would apply logistic regression to determine if the customer will repeat or not.

Step 2 - I will be using transactions data, if needed creating new features on the transaction data summarizing the transaction information for each customer.

I will build a model with limited number of customers as my laptop cannot run models on 28GB of transactions data.

Step 3 – I will try to find an approach where I can use all the 28GB transaction data.

I will be delivering code for EDA, data cleaning, creating new features, applied machine learning models along with project report and presentation deck.