Predicting Customer Loyalty (Kaggle)

By: Vaibhav Prasad Desai, Piyush Bhattad, Abhinav S Panwar, Nishad Gawde

Coupon/ Discounts / Offers are one of the traditional method to attract/acquire new customers at the same time improve relations with existing ones.

The **Acquire valued shoppers challenge** was focused on predicting customer loyalty whether they will become repeat buyers or not based on the coupons/discounts offered to them (redeem).

This dataset gives us a platform to explore into the concept of customer lifetime value but also a Big data challenge where we encounter **350 million rows** i.e the transactional data (anonymized) of over 300000 shoppers.

For executing it we have used cloud technologies: **Google Big Query**(Feature Engineering) and

Microsoft Azure(Machine learning)

These are the four relational files in the dataset

- transaction.csv- contains transaction history for all customers for a period of at least a 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
- offers.csv - contains information about offer

Feature Engineering: “Company”, “Category” and “Brand” represent the most important characteristics

We build features based on We generate a feature: has_bought_company where we count how many times the shopper has bought a product from the company on offer. We generate a related feature: has_bought_company_qwhich holds the quantity bought (sometimes shoppers buy multiple items at once). And another feature that counts the total amount spend on a company on offer: has_bought_company_a.

We also generate features that count the days between the previous purchases and the date of the coupon offer. So if for instance the shopper spend 50$ on a company in the last 3 months we would set has_bought_company_a_90 to 50. We generate these features for last 30 days, last 60, last 90 and last 180 days.

If the shopper has never bought a product from a company on the coupon offer then we generate a negative feature: has_not_bought_company

Same is repeated for Category and brand

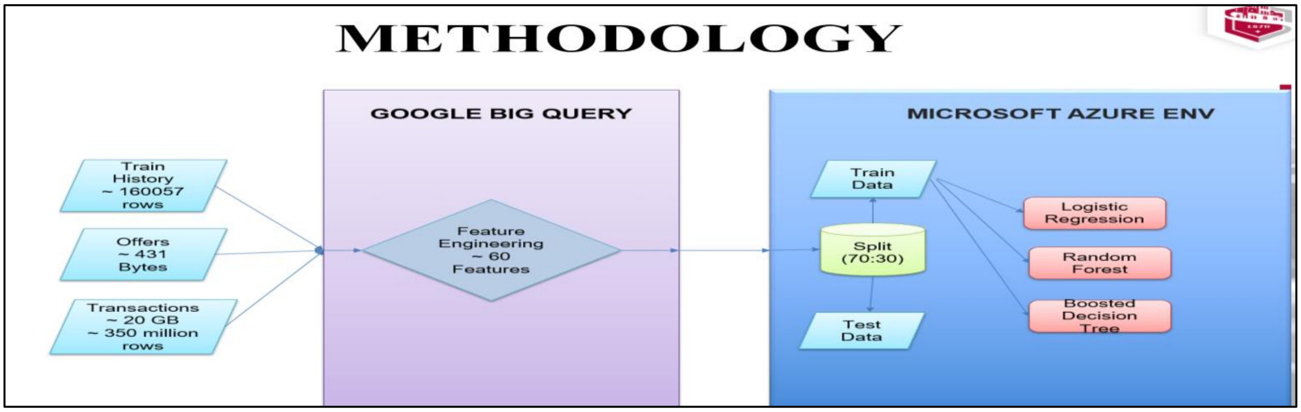
We also build the negative feature for different combinations of Company+category+brand

Along with above RFM(Recency, Frequency and monetary) are included as features for customer.

Total 56 Features including offer value are there

Predicting Customer Loyalty (Kaggle)

By: Vaibhav Prasad Desai, Piyush Bhattad, Abhinav S Panwar, Nishad Gawde



Published Experiment Gallery: https://gallery.cortanaintelligence.com/Experiment/Customer-Acquisition

We do exploratory data Analysis: Variable Type Identification, Univariate Statistics, Histogram Plots
Data Cleaning: We cast column names, Select necessary columns as attributes, and convert "NA" Values =to 0. Response variable is mapped to Binary 1&0

Machine Learning we use below classification algorithms

- Logistic Regression
- Boosted Decision Tree
- Decision Forest

=>Random Split: Train (0.7) & Test (0.3)
=>Response (binary)~Primary features+ Secondary Features

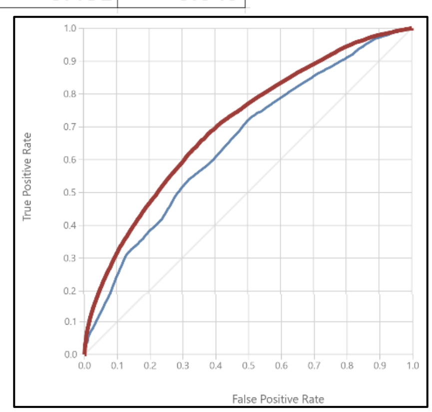
Azure Machine learning gives a great visual interface along with power of cloud computing to experiment with different classification algorithms along with different tuning hyper parameters.

Results for the above three ML methods:

Machine learning Trained models	True +ve	False +ve	False -ve	True -ve	Accuracy	Precision	AUC
Logistic Regression	525	311	12498	34683	0.733	0.628	0.651
Boosted Decision tree	**2886**	**2004**	**10137**	**32990**	**0.747**	**0.59**	**0.704**
Decision Forest	2842	3072	10181	31922	0.724	0.481	0.649

Boosted decision tree has the higher Area under the curve.

We learned that "Feature Engineering" plays an important role in improvement of prediction. Azure ML also gives a platform for data scientists to automate the tedious work of cleaning and data processing along with power of running, tuning, training and, testing different models simultaneously.



Receiver Operating Curve

Reference: https://www.degruyter.com/downloadpdf/j/jaiscr.2016.6.issue-2/jaiscr-2016-0007/jaiscr-2016-0007.pdf

https://mlwave.com/predicting-repeat-buyers-vowpal-wabbit/#comment-289591