Modeling for Marketing Effectiveness: Uplift and Propensity Scoring

Ashwini Rao, Medhaa Bangalore, Shruti Gupta, Srishti Bisen, Yi Chun Ku, Matthew A. Lanham

Purdue University Krannert School of Management

rao120@purdue.edu; mbangalo@purdue.edu; gupta592@purdue.edu; sbisen@purdue.edu; ku19@purdue.edu; lanhamm@purdue.edu

Abstract

Businesses across the world are constantly seeking ways to identify the correct audience for their marketing campaign, and to measure the effectiveness it created. Our project, involves calculating the uplift created by two email marketing campaigns. Uplift models seek to predict, for everyone in some population, the incremental impact of a specific activity on some outcome of interest. Marketing campaigns can be very cost intensive, so identifying the correct audience for every type of campaign is vital. We went ahead to calculate the propensity or the likelihood of a specific campaign making a difference in the conversion of every customer. This helped us identify the ideal set of people to direct this campaign towards to generate maximum profits. Our project can help the marketing team predict the lift for each customer and then empowers them to target only the high-lift customers. Our conclusion is that both campaigns had a positive impact overall, and we've identified subpopulations towards whom these campaigns can be directed.

Introduction

Retailers and marketers seek to identify the returns made on their marketing campaigns. The question this project seeks to answer is how many additional units were sold because of the campaign.

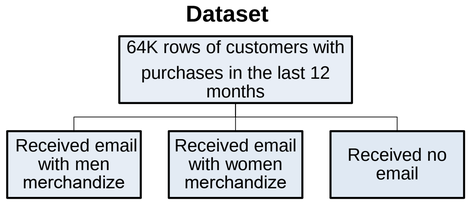

Figure 1. Data Distribution

Uplift modeling was chosen as the predictive modeling technique because it models the incremental effect of treatment on a target group, by forming groups of target and control. This allows marketers to identify customers who respond positively or negatively to their campaigns and those who would have responded positively even when not subjected to a campaign.

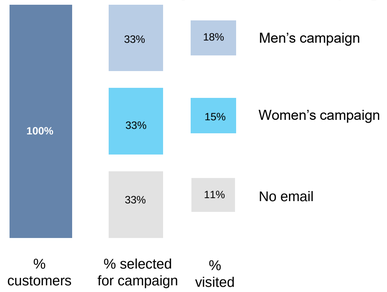

Figure 2. Campaign Distribution

Methodology

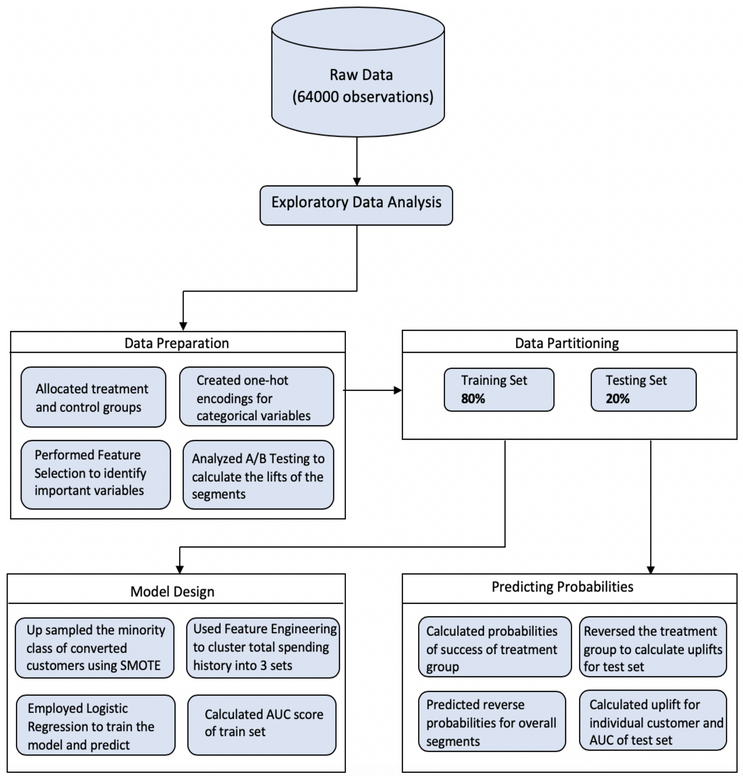

Figure 3. Study Design

Methodology Selection

A propensity model by itself is not going to tell marketers which customers are most likely to contribute to the incremental campaign response. We chose a logistic regression model for the differential response model because it's not only easy to implement and maintain, but is also easy to understand, unlike the black box the neural networks present.

The uplift of a marketing campaign is calculated as the difference in response rate between a treated group and a randomized control group. This allows the marketing team to isolate the effect of the marketing action and measure the effectiveness or otherwise of that individual market in action.

Model Performance Measures

The predictive model was evaluated on its overall accuracy and AUC because identifying the correct Target group of audience is a classification problem, and confusion Matrix statistics help best understand the model performance.

Results

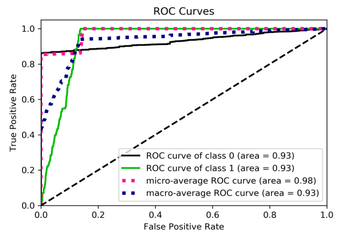

Figure 4. Performance of the model

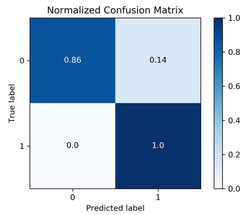

Figure 5. Confusion Matrix

Category	Uplift
Mens E-Mail	0.0083
No E-Mail	-0.0049
Womens E-Mail	0.0069

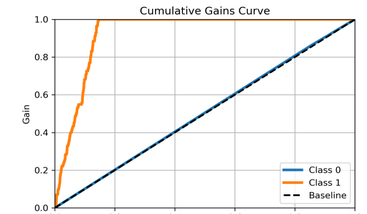

Figure 6. Cumulative Gains Chart

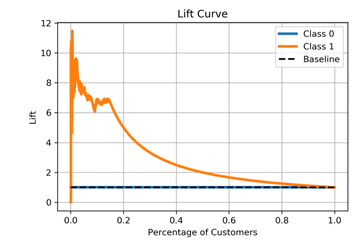

Figure 7. Lift Chart

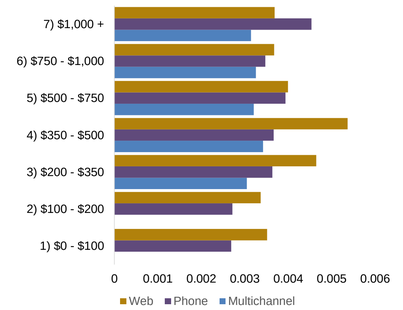

Figure 8. Average Uplift: Purchase `History Vs. Channel

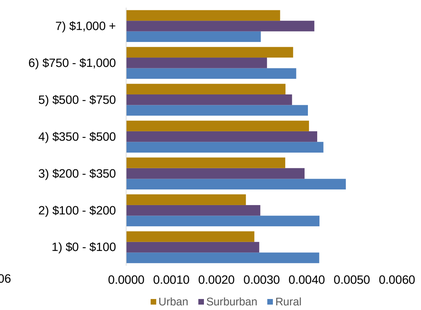

Figure 9. Average Uplift: Purchase History Vs. Zip Code

Conclusions

We observe that the Men's mailing was more effective that the Women's campaign, in terms of increased sales, purchase rate and the site visits.

Targeting top 20% of customers based on propensity would generate best uplift.

The best segments to target are rural and sub-urban web customers with historical spends in the range of $200 to $500

Acknowledgements

We thank Professor Matthew Lanham for constant guidance on this project.