

Duke Datathon 2019: A Valassis Report

Targeted Advertising: An Optimization Problem

Kush Gulati, Ethan Bhojani, Kavi Freyaldenhoven, Alan Suh

Introduction

My eyes snapped open at the distinct sound of my custom ringtone for the Valassis CEO, Katy Perry's *California Girls*. I knew that this 8:30 AM phone call had no chance of having good news. After a late Friday night, I answered the phone groggily: "Good morning Mr. Rubenstein, how are—" that's as far as I went before he interrupted me, "Get to Bostock Library by 9 a.m sharp. There's no time, and we need your help." I knew what this meant, so I assembled a team of adept Duke students and we took the C1 to west campus.

It was here where we were told our task: a covert mission to help Valassis use customer interests and actions to understand what shoppers are most likely to engage with their marketing technology, and ultimately determine the shoppers who are more likely to respond to Valassis ads and convert by signing up, registering, or making a purchase.

With a twelve hour time constraint, we set out to create a robust framework to address this pressing issue. Determining customer conversion is an extremely challenging, yet rewarding task that releases floods of dopamine into our heads, spurring us onwards. In the completion of this project, we cleaned the data and modified it to be readable by the random forests, trained the data, and evaluated it against the validation dataset.

Data Engineering

In the data manipulation and model creation front, the following steps were taken:

- 1) The dictionaries in the second column were spread over multiple columns
- 2) Missing values were filled with zeros

- 3) Columns that differ from the validation and the training data were synchronized
- 4) The most promising features were selected using PCA
- 5) RandomForests with bagging was used to generate the model
- 6) The model was evaluated against the validation set

Findings

Figure 1: Final Random Forced Model results in **0.724 AOC** (100 Trees each w/ depth 6)

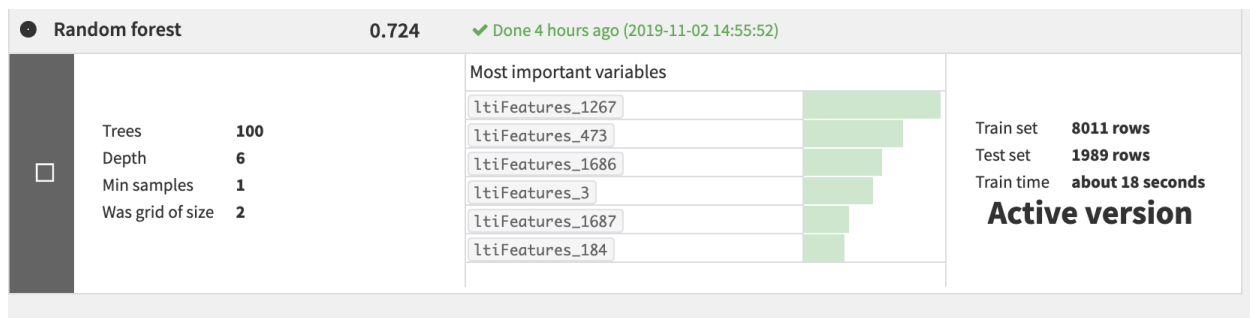


Figure 2: Our Most Valuable Variables (Arts & Entertainment, News, News/Weather)

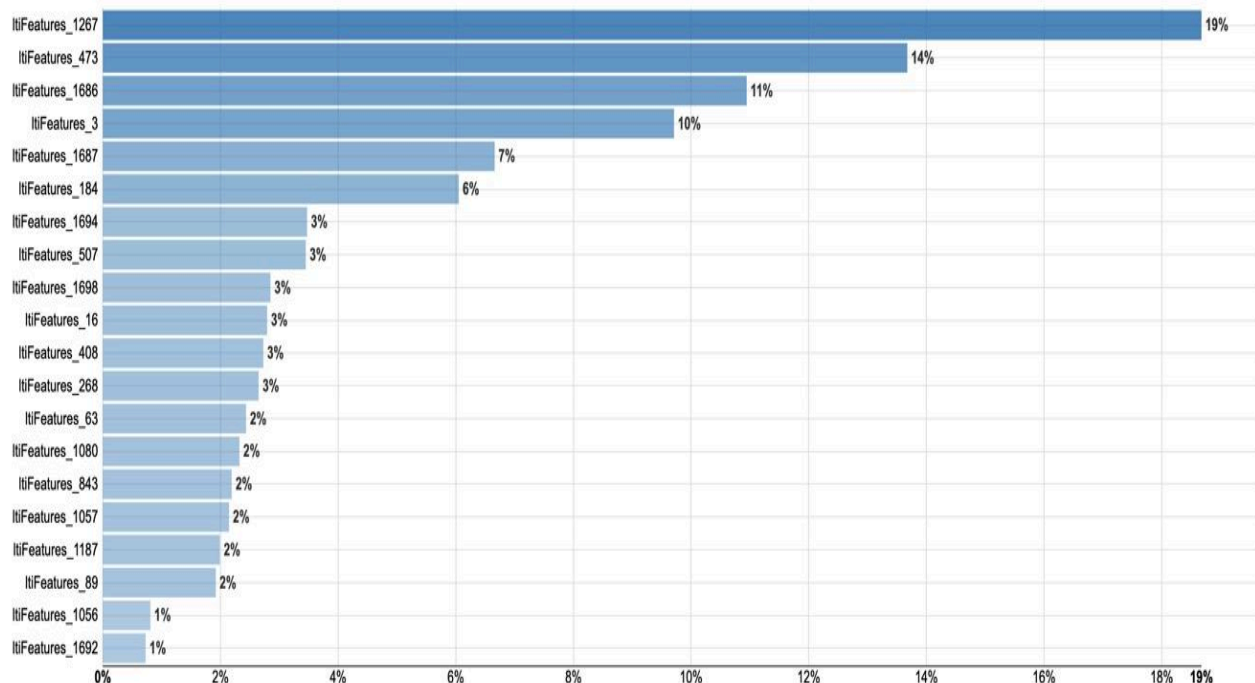


Figure 3: The Receiver Operating Characteristic (or ROC) curve shows the true positive rate vs. the false positive resulting from different cutoffs in the model. The "faster" the curve climbs, the better it is. **The AUC (area under the curve) for this model is 0.724, which is good.**

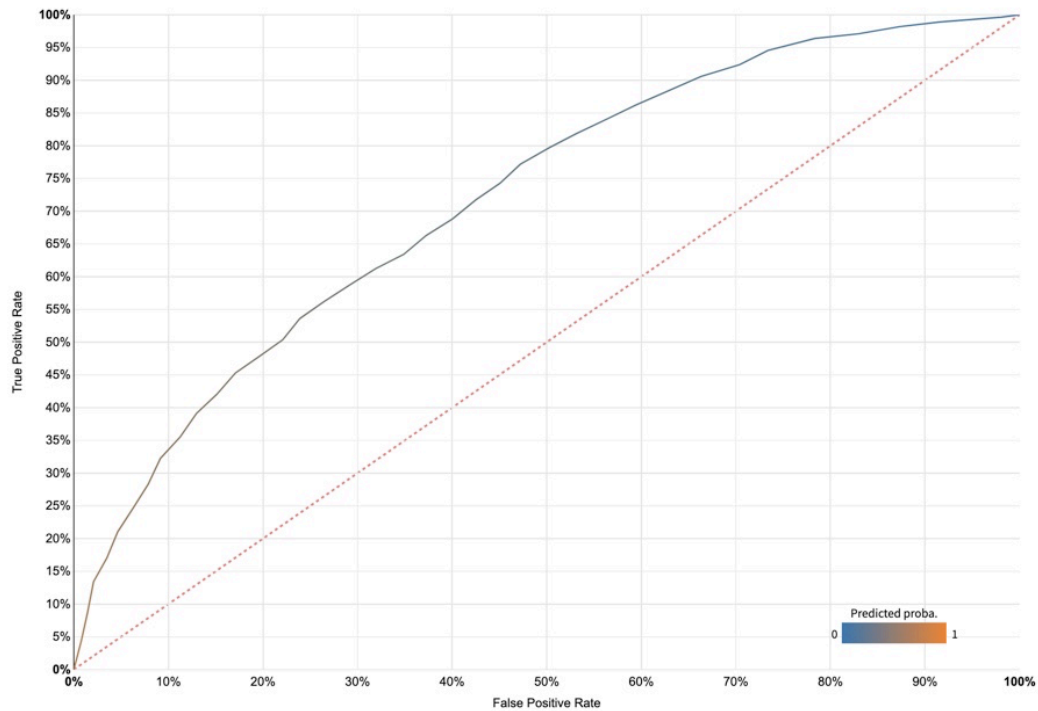
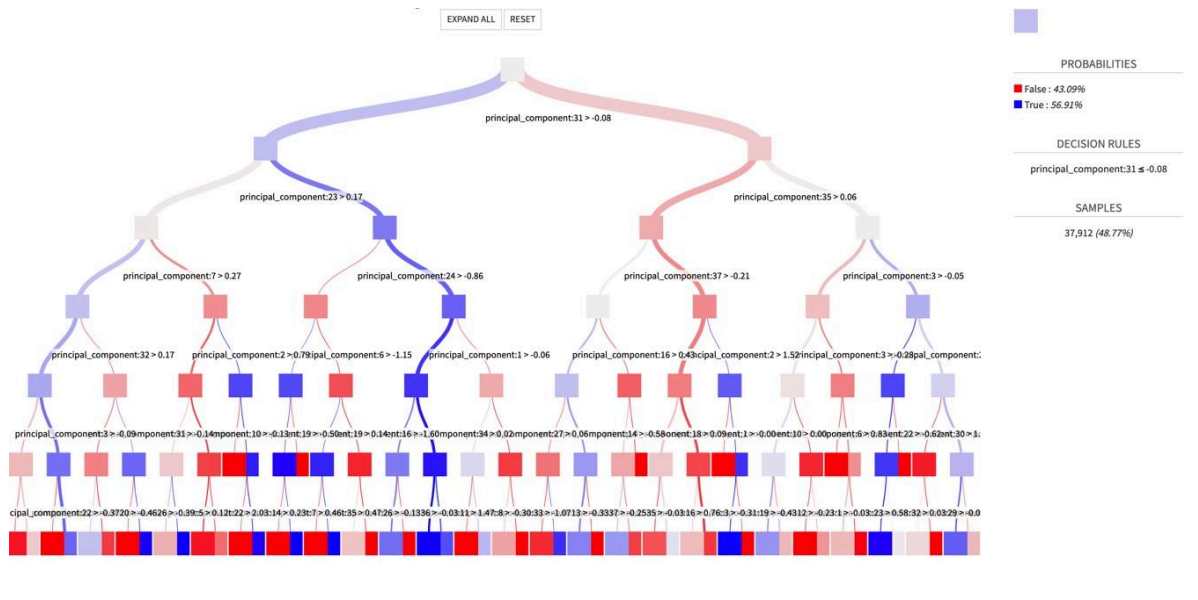


Figure 4: An example of one of our Decision Trees (Expanded View)



Conclusion:

In our model, the most important features are automobile and vehicle preferences as well as arts, entertainment, and news opinions. With this information in mind, we can advise Valassis to advertise to people who have an interest in cars, news, and the arts. We believe this because these features had the most impact and influence on the target variable of our model. We believe that people with these interests are most likely to convert.