Problem: See how accurately we can predict if someone is about to start donating using the commissions

Data Used: Rank 6 months before started donating, Tenure at that rank, Tenure at month, Personal Enrollments

Approach: We created a list of people who are donating using their commissions and then looked at what their data looked like 6 months before. We then created a list of everyone that has never donated using their commissions using their current data on the same data points. This was all put into a single dataset and we attempted to predict which people were going to start donating using commissions.

Problem 1: The vast majority of people never donate, so if you just predict no one will donate you get at 98.9% accuracy on your predictions on the Test Set. So a “highly predictive” result may be no good unless it actually tells us if someone is about to donate.

Accuracy of Model: 98.99% accuracy on Test Set, slightly better than just assuming no one donates. This implies there was some predictive power.

Correlations:

|  |  |
| --- | --- |
|  | **Correlation** |
| **Rank** | 0.23 |
| **PersonalEnrollments** | 0.20 |
| **TenureAtRank** | 0.08 |
| **TenureAtMonths** | 0.02 |

Size of Data Set: 391,879 distributors

Creating a Predictive Set:

What if we just took only those that the model predicted would donate and looked at only those?

Imagine a Drip campaign where we took only the most likely people to donate according to the model we trained and only wrote to them. There were 391,879 people that were in the original set, but the model predicted only 1167 of them would donate and did so with over 65% accuracy. Imagine a Drip campaign that had a 65% success rate! This is unrealistic because these are people we know donated, but it’s suggestive of what is possible.

If a Drip campaign was normally successful 3% of the time and instead we doubled it to 6% what would be a doubling of the size of donations to the foundation. So we don’t have to be very predictive to be a powerful tool for the foundation.

I tried adjusting the tolerances on the model to see if I could get it to give me people more likely to donate even if the change it predicted was only 10%. This adjusted the number of predictions to 11329 people that it predicted would donate, but dropped the prediction rate to 25%. This may not sound good at first, but imagine a Drip campaign that was 25% successful. These are actually very high numbers.

Recommendations:

We believe the Foundation could benefit from Machine Learning and Predictive Analytics. These results could probably be improved quite a bit with more effort. But our initial analysis suggests that there is strong predictive power in the data that would allow the Foundation to create a list of people likely to start donating and advertise the benefits of the foundation directly to them. This seems likely to increase donations by a fair margin.

The next step would be to try this out and then collect real data on how many people start to donate from a drip campaign and how this increases Foundation donations.

There are other possibilities here as well. We could imagine a change to the Virtual Office that prompted people likely to want to donate to round if they didn’t choose to. This would keep the interface from being ‘annoying’ to those that aren’t likely to donate, but prompt those that find that a useful prompt.

Presentation:

1. Who is the team? What are our skills?
2. What were we trying to do: predict if someone would donate in the next 6 months
3. Why? Because we wanted to see if it was possible to predict future donations
   1. Possible applications:
   2. Drip Campaign – find high probability candidates
   3. Determine if someone wants to permanently round
   4. Determine who responds to email campaigns
   5. When do people decide to start donating? i.e. when reaching certain ranks? Doing certain enrollments?
4. Correlations with data
5. About the model
   1. Only 4 features
   2. Using Logistic Regression (simplest possible ML model)
6. Numbers:
   1. 98.9% of commission earners don’t donate via commissions (392,000 in our set)
   2. Our model predicted 1167 would donate, of which 65% actually did
   3. If our model is asked to predict anyone with even a 10% chance of donating, this number jumps to 25% of 11329
7. Future Directions
   1. Improve data: use 100s of features instead of 4
   2. Improve model: Deep Learning (Artificial Neural Networks), k-Nearest Neighbors, Support Vector Machines, etc. Or use ensemble of all of them to improve model.
   3. Create email campaign and then track who donates and who doesn’t. This would be much better data than using commissions donations as a proxy.
   4. Predicting how much people will donate to focus on potential high donners.

Ideas:

* Get a picture of Logistic Regression
* Explain about Deep Learning’s recent amazing successes, i.e. Facebook recognizing faces