Donors Choose: Identifying Projects at Risk of Not Being Fully Funded

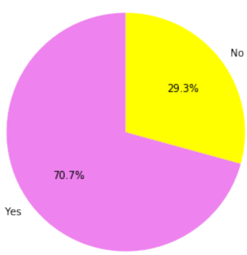
Mario Moreno

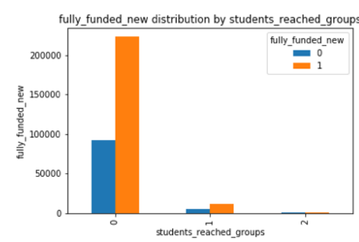
Introduction

At any given moment, thousands of teachers in K-12 schools propose projects and funding requests on DonorsChoose.org to enhance the educational experience of their students. Some of these projects receive full funding from DonorsChoose.org's vibrant user community, yet other proposals go unfunded. In an effort to support important projects across the educational spectrum, this analysis seeks to help DonorsChoose.org better predict which projects might be destined to miss their funding goals in order to provide them the help they need to make their project proposals more attractive to prospective donors.

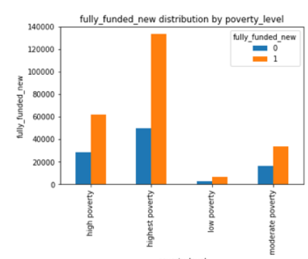
Data Exploration

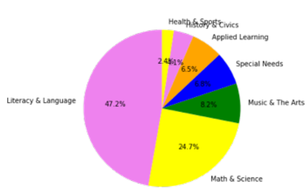
In order to determine what characteristics successful projects shared, we matched the project descriptions in the projects dataset with the outcomes dataset for the 2011-2013 timeframe. This gave us an initial set of features with which to work. Using this dataset as a starting point, we explored the data to determine a few overarching characteristics of DonorsChoose.org projects.





In general, a majority (70.7%) of projects posted within our timeframe were fully funded and most of that funding financed targeted projects that helped small groups of students rather than large groups of students. Additionally, the greater the poverty of the beneficiaries of a project, the more likely DonorsChoose.org funders would fully fund a project.





A vast majority of the project proposals requested funds for Literacy and Language projects, which corresponds with DonorsChoose.org funders' choice to overwhelmingly fully fund projects in this subject area. Projects in mathematics and science were the second most-requested subject areas, and were similarly more likely to receive funding when compared to other subjects.

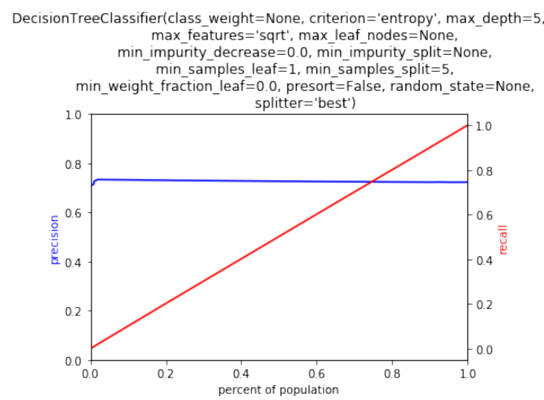
Building Features

In order to carry out a machine learning analysis to help better predict whether a project would be fully funded, we needed to develop a set of characteristics that could help us make that prediction. Starting from our original merged dataset, we began digging into the features within that dataset. We explored correlation between features, removed those that were randomly assigned (such as projectid), discretized continuous variables such as age, imputed for missing values, and dummified a number of variables. We ended up with 118 features for our machine learning analysis.

Results

To predict which projects would be fully funded or not, we deployed 60 models that utilized different iterations of parameters for 6 different classifiers. In each of these models, we wanted to determine whether the outcome would be positive (fully funded) or negative (not fully funded) within the given timeframe.

In measuring the overall accuracy of the models, we prioritized the AUC-ROC score to determine how much better our models perform over random classification of positive or negative scores. In analyzing this metric, the majority of the AUC-ROC scores come in at roughly 0.5, indicating that these models are barely better than random assignment of outcome values, if at all. That said, the model with the highest AUC-ROC score is a Decision Tree model split on entropy, with a max depth of 5, sqrt max features and a minimum sample split of 5. The precision-recall curve for this model is on the right.



In fact, across all relevant metrics, Decision Tree models outperform other models on average. Decision Tree models are the most precise on average at 10 and 30 percent of the population, and they have the highest recall at 10 and 30 percent of the population. For lower levels of precision and recall, AdaBoost works the best on average.

Most of our precision-recall curves looked like the one pictured above. Overall, our analysis performed well on precision, consistently delivering results at around 80% precision regardless of the percentage of population. Our analysis did not perform well in recall, but did improve for smaller subsets of the population. This suggests that our models care more about casting a wide net from which to choose projects that were not fully funded, rather than valuing recall.

Recommendation

The results of our machine learning algorithms are generally weak, which is why we recommend that DonorsChoose.org build upon our current work to strengthen the predictive capacity of our model. In particular, DonorsChoose.org should endeavor to improve upon the features we've selected in our analysis in order to increase the predictive power of the machine learning methods we test across.

Once a richer set of features is built, we suggest that DonorsChoose.org deploy a Decision Tree model first to confirm that it's the strongest of the models. From there, we suggest that DonorsChoose.org expand the timeframe to encompass the entirety of their dataset in order to collect more robust results.