James Jensen
Machine Learning for Public Policy
Assignment #3: Improving the Pipeline

DonorsChoose: An Analysis of Unfunded Projects

Background

DonorsChoose is the largest non-profit in the U.S. dedicated to connecting individual donors with school teachers in need. The website provides a place for teachers to post a description of their project, including the reason why they need resources, the number of students it will serve, and the total cost. Potential donors have the opportunity to learn more about each project, and in doing so decide who they wish to financially support. With dramatic cuts to school budgets, and with over half of all projects listed as 'highest_poverty', it's clear that they're not only serving classrooms across the country, but those which are most at need. However, currently only 71% of all projects on DonorsChoose receive funding within 60 days. The analysis I've undertaken aims to predict if a project on DonorsChoose.org will not get fully funded within 60 days of posting. By better understanding the shared characteristics of projects that do not receive funding, DonorsChoose can help teachers maximize their efforts and donor money by providing actionable ways to improve their initial request as well as how they market their project.

Analysis

I compare the performance of 7 different types of models -- logistic regression, k-nearest neighbors, support vector machine and decision trees (simple classifiers), as well as bagging, boosting, and random forests (ensemble methods) -- across a host of parameters to best understand what model provides us with the results that are most aligned with question we're interested in answering. I also train and test the models over multiple temporal splits with a rolling period of 6 months -- the size of both training and testing sets are 4 months, with a 60 day gap between the end date of each training set and the start date of each testing set.

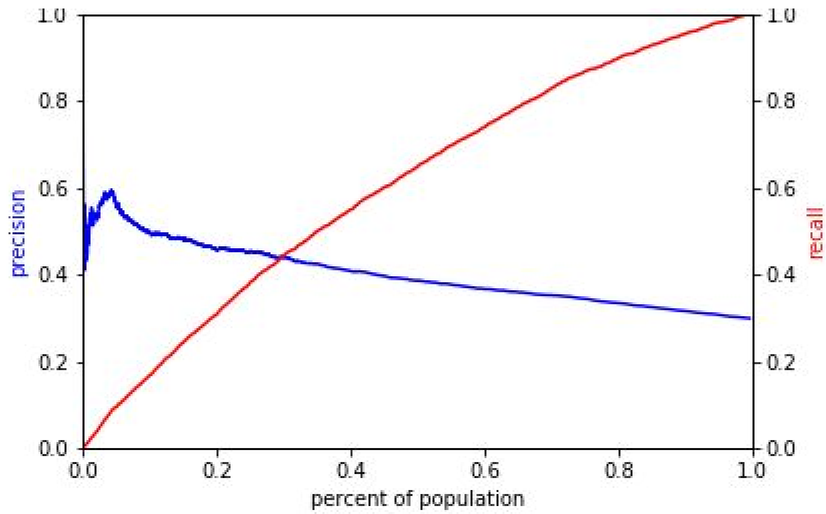
Given that most of the highest performing models have an AUC between .6 and .7, it's clear that the models are slightly better than random guessing but none are particularly strong predictors of projects that won't receive funding within 60 days. If we're interested in balancing both precision -- in this case, the number of actual projects that did not receive funding within 60 days over the total number of predicted unfunded projects -- and recall -- the ability to predict all cases that will not receive funding, then we should choose the model with the highest F1 score. We can see that across the top performing models there is a massive jump in recall when shifting the threshold from .1 to .5 without a comparable decline in precision, suggesting that increasing the threshold can increase overall performance. This bears out in the F1 score -- among the models with the highest AUC (logistic regression and bagging) -- we see it increase by .1 points.

Across all time periods, there is less variability in performance among logistic regression and bagging than random forests and decision trees, suggesting that the bagging classifier is the top model. However, even for this model the accuracy is not significantly distinguishable from our baseline metric. With a baseline of .28 (which represents the total percentage of projects that were not funded within 60 days out of the entire set of projects) and accuracy score of .72, it's possible that the model is classifying each project to be funded within 60 days and misclassifying all projects that we're interested in -- the ones that remain unfunded.

Recommendation

For someone focusing on the top 5% of posted projects that are at highest risk of not getting funding, I'd recommend using bagging classifier (trained from 2012-01-01 until 2012-11-02 with 20 estimators), since it is the model with the highest precision at 5%. Given that we're focused on a limited sample with the highest risk, we want the intervention to guarantee that it's prioritizing the projects at need. By optimizing for precision, of the projects we identify as 'not funded within 60 days', we have the highest likelihood of intervening on projects most at need of intervention. With that said, there is not a noticeable difference between the accuracy score for this model and the baseline metric. Therefore, before these predictions are used to inform intervention, it is recommended that additional modeling occur.

Precision - Recall Curve for Top Performing Model



As you can see in the plot above, the bagging classifier (trained from 2012-01-01 until 2012-11-02, with 20 estimators) had precision at top 5% of the population close to .6, which was the highest recorded value at this threshold. However, when increasing the threshold to intervene on a larger percentage of the population, precision quickly declines. This drop-off in precision confirms that this model can be useful when identifying those most at risk of not receiving funds, but should not extend to a larger population.