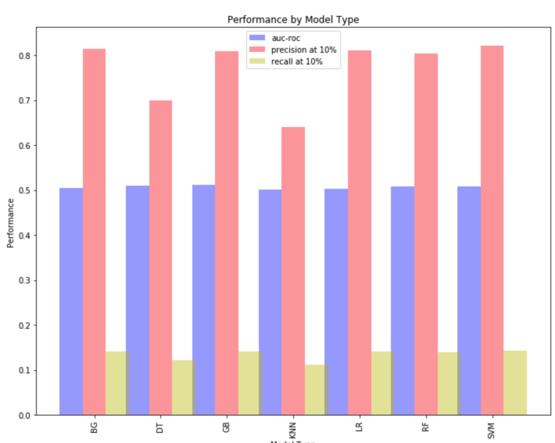
Donors Choose: Predicting Projects at Risk of Not Being Fully Funded

I. Model Training Results

To achieve the goal of predicting which projects will succeed or fail at the time of posting, we tested 222 different models, including 7 different classifiers and a variety of different parameter values for each model. In all of these models, our outcome of interest was whether a project would be fully funded, with full-funding classified as a positive result, and not full-funding classified as a negative result. The goal of all of our models is to successfully assign projects at the time of posting to one of these two outcome classes. To assess the overall accuracy of our models in the entire space, we use the auc-roc metric, which is equal to the probability that a classifier will assign a randomly chosen positive observation a higher likelihood of being positive than a randomly chosen negative one. In reviewing the results of our model testing, we see that none of the models performs very well in terms of overall accuracy, hovering at around .5 auc-roc across all models. This means that most of our models perform roughly on-par with randomly assigning models into classes (noting that the random baseline accuracy for our largest temporal split is .512).

The model with the highest auc-roc at .523 is the Random Forest with maximum depth of 20[1] on the largest temporal split[2]. Based on the initial results, we would **recommend that Donors Choose use this Random Forests model.** However, given the weak overall performance of all the models and the very similar performance across models as shown below, we recommend additional analysis (both generating new features and eliminating useless features) to improve overall performance. Indeed, we would caution Donors Choose against using this model to generate whole-of-population predictions. That said, the models do perform well on the top segment of the population and offer useful lessons learned regarding which project features are most correlated with success or failure in securing full funding, discussed further below.



Based upon the results in the full outcomes table, we see that performance is fairly consistent over time, increasing slightly as the training set grows from one month to three months. This is an encouraging finding, as it suggests that the model Donors Choose ultimately uses will perform consistently over time – reducing the need for highly frequent re-training of the model – and that having more data in the training set yields a model that is less overfitted to the specific training data yielding a slight improvement in the accuracy on the testing data. As the amount of training data that Donors Choose has will grow over time, the strength of the predictions may also improve.[3]

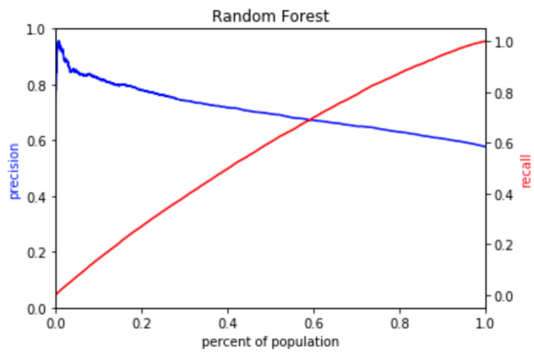
---

[1] The full set of paramters is as follows {'max_depth': 20, 'max_features': 'log2', 'min_samples_split': 10, 'n_estimators': 100, 'n_jobs': -1}

[2] Training set: January-March 2011, test set: April 2011.

[3] While we performed the main model testing on a subset of the data (January-April, 2011) due to extremely high training times on the full data, as a validation we also trained the Logistic Regression and Random Forests models on the full time series with a 6-month temporal validation split. These two models were chosen because of their strong performance on the shorter time-series and fast training speed. We see that both models perform best on the longest training set, with results fairly consistent to the shorter temporal splits (logistic regression achieving a maximum auc-roc of .5398 and random forests achieving a maximum auc-roc of .5246).

Looking across all of our models for the largest temporal split, we see generally consisting performance, noting that k-Nearest Neighbors performs notably worse than the other models. This may be caused by the fact that k-Nearest Neighbors is particularly sensitive to the inclusion of irrelevant features, as this model considers all features included in the model when calculating the distance between observations. This would suggest that the model could benefit from pruning some of the irrelevant features and re-running the model.



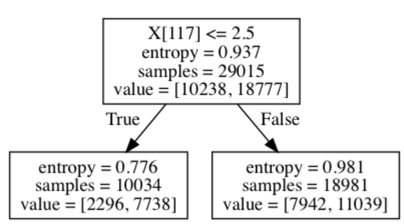
It is worth noting that the models performed far better on precision (the ratio of true positives to all predicted positives) than recall (the ratio of true positives to all observed positives), with average precision for the better performing models maintaining scores of above 80% at 10% of the population. The fact that precision performs better than recall suggests that our models may be over-classifying projects as not-fully funded which avoids false positives (preserving precision) but limiting both recall and accuracy. We show the precision-recall curve for this best performing model at left.

II. Recommendations for Donors Choose

feature_import	cols
0.201741	total_price_including_optional_support_bins
0.184183	total_price_excluding_optional_support_bins
0.161894	eligible_double_your_impact_match
0.072011	resource_type_Technology
0.056136	resource_type_Books
0.029287	school_charter_ready_promise
0.027695	poverty_level_highest poverty
0.027472	teacher_teach_for_america
0.026907	school_metro_urban
0.022204	eligible_almost_home_match

To understand which features are most strongly correlated with whether a project will be fully funded at the time of posting, we identified the most important features[4] in our best-performing Random Forests model. The table below shows the top 10 most important features. We see that price is the most important feature, with total price including and excluding optional support as the first and second-most important features.

Using a simple, one-level decision tree below, we verify our finding that total price offers the greatest gain in classification (measured in the reduction in entropy, which is roughly the mix of true and false observations in each category). This suggests that, in general, projects in the lowest 20% ($0-$300) of pricing are more likely to be fully funded, and conversely projects above $300 threshold are at significantly higher risk of not being fully funded. Indeed, when we look at the proportion of projects that are fully funded by price between 2011-2013 across the full dataset below, we see that the proportion of fully funded projects consistently decreases as the project price increases. We therefore would suggest that Donors Choose advise teachers to **limit the price of their projects (or split large projects into multiple, smaller projects)** to have the best chance of full funding.



total_price_bin	fully_funded
0	0.895691
1	0.822275
2	0.780301
3	0.730421
4	0.697715
5	0.675594
6	0.666062
7	0.619360
8	0.581712
9	0.544726

[4] I identified the most important features based upon the sklearn feature_importances_ method.

To better understand the direction of the correlation between other features in the top 10 most important features and the funding outcome for a project, we plot both the variable averages by whether a project is fully funded along with the correlation matrix for the features and outcome variable. We see the following results:

- Eligibility for double your impact match is positively correlated with a project being fully funded (or negatively correlated with a project not being fully funded). Therefore, one intervention that Donors Choose could take for projects at risk of not being fully funded is **to make those projects eligible for double your impact match**. Almost home match also increases the likelihood of a project being fully funded, though its effect is weaker than double your impact match.
- We interestingly see that the technology resource type is negatively correlated with full funding while the resource type of books is positively correlated with full funding. This may suggest that donors prefer to fund books over technology resources. Therefore, **technology projects may be at high risk of not being fully funded**.
- We see that the following characteristics are weakly positively correlated with full funding: schools in urban areas, highest poverty schools, Teach for America teachers, and charter ready promise schools. We would recommend that Donors Choose **encourage at-risk projects that fall in these categories to emphasize these factors in their posts** to improve chances of full funding.