HOMEWORK 5 – POLICY MEMO WRITE-UP

Allison Collins

BACKGROUND & GOALS:

The goal is to allocate interventions for DonorsChoose projects by predicting whether or not a given project will be funded within a 60-day period. I tested a variety of models on a dataset spanning the year 2012, based on a newly revised pipeline. A separate Jupyter Notebook houses running the code and analysis and select outputs for running each of the models (which can also be seen separately in a csv file, with precision-recall curves; printed trees; and feature lists all also stored separately in the repository); this write-up serves to primarily discuss policy implications rather than technical explanations.

MODEL GENERATION & EVALUATION:

As can be seen in the output csv, performance varied across the different types of models, as well as within a given model type based on parameter choice (precision, AUC, recall). Looking at area under the curve, the highest performing model was a Random Forest model built on the test/train dataset with start date for testing data of May 1st. The AUC was .999. Conversely, models that performed poorly included the worst AUC seen with a bagging classifier, built on the same date (May) dataset with AUC of .48. Some of the gradient boosting models with higher number of estimators performed very well on precision, but not quite as well on area under the curve (more in the range of .7). Logistic regression also saw comparatively lower levels of precision, and decision tree performance varied across the models, with some with higher depth performing more strongly. As expected, precision decreases as we raise the k threshold—and recall increases, across the models.

Overall, random forest models with higher depth performed strongly on precision at the top 5 and 10% (those with lower depth saw substantially lower precision), with several of the models having precision of 1 at these high thresholds (and were strong on AUC). Looking over time in the aggregate, I do not see a much higher performance for different dates – for example, in the Jupyter notebook, we can see that the recall improved slightly at both 5 and 50 over time, but overall there does not seem to be that much of a difference. The later dates get a little more data to work with, so it would make sense for them to perform slightly better, but I did not see that much change when investigating the results csv and checking via some quick plots.

When investigating general performance (compared to what we have learned to be "good" scores, as many of the models do not do that much better, say looking at a "baseline" AUC of 50%), resources suggested that feature selection is key, but a model with the right feature selection when stacked against a different model (e.g. tree + random forest) using the same features – on the whole, recognizing models fit different problems/datasets better – will perform more similarly than the same model type (e.g. decision tree) using a bad set of features and a good set. Notably here, the project's total price is a strong feature, which has implications when considering how to allocate among projects. Also overall, baseline accuracy is ~33% and thus we can see improvements from using one of these models.

One last note on evaluation in seeking to find the “best” model to use is that in some cases, we may be faced with the tradeoff between a more understandable model and one that better predicts outcome—for instance, it would likely make the most sense here to use a Random Forest, if we are trying to allocate scarce resources to help bolster the top projects (and focused on correctly capturing), but this may be harder to explain to a policy audience than say a decision tree that does almost as well.

Prioritization of projects for Intervention:

The choice of which model should be selected to inform predictions given that only 5% of the projects can be intervened on greatly depends on the priorities of those who are planning the intervention. The models suggest ways to predict which projects will not be fully funded in the 60-day time period. However, deciding how to spend funding requires answering several strategic questions, which I will tee up and present ways that we could think about making choices, were we to have more information on how to answer these:

- What is the policymaker’s priority – is it to provide assistance to the projects that have the least chance of receiving funding, or to “push projects over the edge” which may get close?
- How is equity being incorporated? Is there consideration being given to whether there are certain features of a school (for instance, being low resource or high poverty) that will additionally be considered
- How will resources be split and what is the funding amount for these interventions?
- Are we more concerned with missing projects (e.g. we are afraid of false positives, where here a positive 1 means getting funding) or with avoiding funding projects which actually would be likely to be successful on its own (here, a false negative)?

With a deeper understanding of the specific goals at hand, we could take specific steps to identify the optimal 5% of projects to support

- If we are concerned with maximizing the correct prediction of true positives (here, correctly predicting the highest amount of positive projects not being funded and incorrectly classifying ones that won’t get funded as getting funded), to make sure we are not missing out, then we should pick the model with the highest recall
- If we are concerned with the highest amount of positives right (e.g. models that will not be funded are correctly predicted as such and we are not falsely predicting some to not get funded), we should pick the model with highest precision
- If it relates to projects at the margin vs. lowest probability, we could look at relative ranking of scores and how close to the threshold cutoffs

In the case of this dataset in particular, we must consider that the project’s total price is a key feature, and whether while that is a strong predictor, it is the way in which we want to structure our intervention (again returning to the question of priorities and equity).