Vedika Ahuja

5/2/2019

HW 3 Report

Machine Learning Policy Memo

**To:** DonorsChoose.org

**Subject:** Supervised ML Models Used to Predict Donor Projects Least Likely to Receive Funding

**The Evaluation Metrics of the Models**

*Precision*

The models with the highest precision at thresholds between 1 and 5% are Support Vector Machine (SVM) and K-nearest neighbors (KNN). These models do significantly better than Random Forest (RF), Decision Tree (DT), and Adaptive Boosting. The Random Forest model’s precision increases as the threshold goes to 10%, and is actually higher than the KNN model at 10% and beyond. Decision trees, with and without boosting, both have the lowest precision at thresholds 30% and smaller.

In other words, when the models rank each project from most likely to be funded to least likely to be funded, the projects that the SVM and KNN models rank as least likely to be funded are more likely to be correctly classified than the other models. For example, when looking at 1% of the projects posted that are the least likely to be funded ranked by the KNN model, between 34% and 43% (depending on the validation period) of those projects were actually not funded in the validation set. For the SVM model, the number is between 35% and 47%. In comparison the decision trees, random forests, and adaptive boosting of decision trees have almost a 0% precision rate at the 1% threshold.

*Recall*

Models with highest recall at thresholds between 1 and 5 are the KNN and SVM model again, and the Random Forest model has a comparable recall at the 10% threshold level and beyond. The Decision Tree with and without boosting have significantly lower recall levels at thresholds below 30%. Recall tells us how many of the projects that failed are actually identified. So at lower threshold levels, recall will be lower, because there are inherently fewer projects chosen. In a policy sense, recall tells us “coverage.” If someone wanted to intervene on the most possible projects at risk of not being funded, then they would want a model with a high recall score.

*AUC Score*

The SVM model had the highest AUC-ROC score, followed by the KNN model, the Adaptive boosting of the DT model, and the DT model, all of which had very similar AUC-ROC curves. At a high level, the AUC-ROC score is an overall performance measure for how the models classifies the projects at all threshold settings – it measures how the model can distinguish between the classes. Given the policy problem stated, this score is not necessarily the most helpful in picking the best model to design an intervention around.

**How the results change over time**

I would expect that precision and recall improve as the training set grows, so when the date for splitting the training and validation set is later. However, this does not seem to be the case. With the KNN model, the validation date (I will define this as the date used to split the training and test sets) with the highest precision and recall rate is different at different thresholds. In the SVM model, precision and recall actually consistently decrease with later validation dates for all thresholds. This could mean there is a time effect on the chances of a project getting funded.

I would recommend for a person working on this model who wants to identify 5% of the posted projects that are at highest risk of not getting fully funded to choose the \_\_\_ (K nearest neighbors or SVM) model. This is because the evaluation metric of importance to the person is precision, and precision specifically at the 5% threshold level. The person wants to intervene on the projects with a probability risk score at the 95th percentile or higher, and so we want to pick the model with the most true positives beyond the 95th percentile. In other words, the person doesn’t care about correctly identifying ALL the risky projects. Instead, they want the model to correctly identify the riskiest projects.

**Conclusion**

It is important to compare the models with baseline, which is the rate you will correctly identify a project that won’t get funded by choosing the project randomly. That number is approximately 30% (ranges depending on the validation set). The precision rate of the best model at the 5% threshold level is around 42%. This means that if you pick randomly from the top 5% of the projects, ranked by the model from most to least likely to get funded, then 42% of those projects were actually not funded in the validation set. This is substantial difference, which means even the best model I created doesn’t substantially help the person trying to help the project most at risk.

Note: I was not able to run the SVM model on the full sample, so I also ran all the models on a randomly selected quarter of the sample. The information above is using results from both outputs.