Report of evaluation of different models on predicting the failure to get grant in 60 days

Peng Wei

I. Model pools & Training Testing data & Parameters

We treat the failure of getting fund in 60 days as a positive result, a success to get grant in 60 days as a negative result. To choose the best model, we construct a model pool that contains the following models: Random Forest Classifier, Extra Trees Classifier, Ada Boost Classifier, Logistic Regression, Support vector machine, Gradient Boosting Classifier, Decision Tree Classifier, Bagging Classifier. All models are built with different parameters (*appendix 1*), different training & testing data (*appendix 2*).

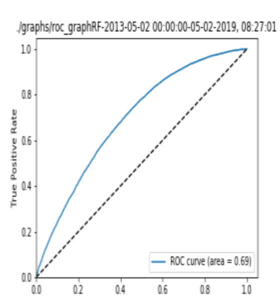
II. Model performance

1. Best metrics that can serve our policy goal

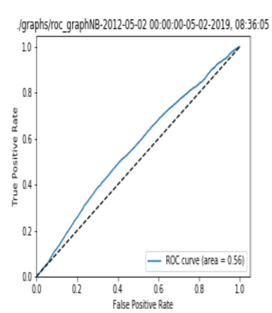
For the model, we care about two aspects:

The general performance that it can distinguish the failure and the success. For this purpose, AUC-ROC is a good metric to use. It's standardized covering curve. Generally speaking, the area under the curve represent how well the model can predict the outcome right. *Graph 1 vs Graph 2*, it's clearly that model of graph 1 is better than model of graph

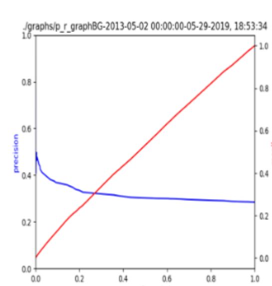
With the budget constraint, we want to focus on the projects that can't get funded in 60 days. So we can help them to find the funding. We want to maximize the share of the true positive (not get grant in 60 days) in all predicted positive. Therefore, I recommend we focus on precision. We can use the precision- recall curve with percent of population as x-axis. When we set the threshold to k, it means that if we have the resource to provide top k percentage of the population, which model give us the bigger precision and recall we would like to choose it as better one. Graph3 vs Graph4, it's clearly that model of graph 3 is better than model of graph 4 when k = 20.



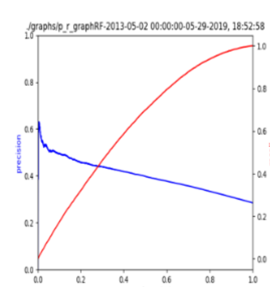
(Graph 1)



(Graph 2)



(Graph 3)



(*Graph 4*)

2. Same model's change over time (use different training data and testing data):

Among the models that share the same parameters, trained from different data set, there isn't much differences. Some model will performance a little better with larger training data set. It makes sense that we always want more good data to build model. The other key point is that all the models doesn't change much, after deploying it, we can expect it will work consistently.

3. Which model performs best under the condition that k = 5

For the testing grid (21 models):

Random Forest performs best under the condition of k =5. As mentioned before, the donor cares about two metrics, one is AUC-ROC, the other is precision. Although, Bagging performs better when k =5, in term of precision.

However, after we check the ROC, Random Forest is better than Bagging Meanwhile, it performs very stable over the time. It's the best model that we can use if we set the threshold to 5.

For the small grid (186 models):

LR performs best under the condition of k =5. It performs better when k =5, in term of precision. Its auc-roc is 0.653305, which is not far away from the auc-roc score of the best model. We can use LR model for the project if the budget is enough for 5% of the whole project.

III. Observations

It's not surprising that with the same training, testing data, the embedding models generally perform better than the common models. Random Forest is better for almost all metrics.

For baseline, we use dummy classier with method of stratified to do prediction at every threshold of K. I choose precision for dummy classier. Most of time, the other models perform better than the baseline. However, It seems that KNN is not as good as the dummy classier in some cases.

IV. Recommendation for donors

Since the donors have budget constraint, they want their money to have a good affect to the school as soon as possible. I would recommend they use the LR as their model to deploy.

Appendix 1:

Training & Testing data contains the following

ID	Traing data	Testing data
1	2012-01-01 to 2013-05-02	2013-07-01 to 2014-01-01
2	2012-01-01 to 2012-11-02	2013-01-01 to 2013-07-01
3	2012-01-01 to 2012-05-02	2012-07-01 to 2013-01-01

Appendix 2:

Table of the best models in term of different metrics

model_type	metrics	train_end	score
43	AB	auc-roc	2013-05-02
26	LR	accuracy_at_1	2012-05-02
35	RF	accuracy_at_2	2012-05-02
29	RF	accuracy_at_5	2012-05-02
40	AB	accuracy_at_10	2012-05-02
45	AB	accuracy_at_20	2012-05-02
35	RF	accuracy_at_30	2013-05-02
12	DT	accuracy_at_50	2012-11-02
35	RF	precision_at_1	2012-11-02
35	RF	precision_at_2	2012-11-02
24	LR	precision_at_5	2012-11-02
43	AB	precision_at_10	2012-11-02
43	AB	precision_at_20	2012-11-02
39	AB	precision_at_30	2012-11-02
12	DT	precision_at_50	2012-11-02
35	RF	recall_at_1	2012-11-02
35	RF	recall_at_2	2012-11-02
24	LR	recall_at_5	2012-11-02
45	AB	recall_at_10	2013-05-02
35	RF	recall_at_20	2013-05-02
35	RF	recall_at_30	2013-05-02
12	DT	recall_at_50	2012-05-02