Quinn Underriner
Policy Memo, Machine Learning for Public Policy HW4

Introduction

DonorsChoose is a US based nonprofit that allows teachers the ability to directly appeal to the public for donations for their public school classroom needs. The United States has high variance in how much it spends on public education as in many communities this funding is based on property taxes.[1] These leads to many severely underfunded communities. Further, the amount that the United States invests in education is declining.[2] In the absence of increased funding, better targeting of these types of charity will be more crucial to the educational system. Currently, 29 percent of projects posted on DonorsChoose do not get funding within 60 days. This analysis intends to predict which projects are not likely to get funded. With this analysis, policy makers can target interventions with projects that are not likely to get funded to help ensure they do. Further, having identified what are the attributes of good project proposals, DonorsChoose can provide great guidance to teachers before they put sites together, potentially saving them time on preparing their pleas (leaving them more time for teaching).

Result From Training the Model:

After applying 7 different models with a myriad of potential parameters, a bagging model (parameter, n estimators=20), has the best f1 score, which means that of the models it did the best job of minimizing false positives and false negatives. This model has a solid precision rate of 40 percent (at a threshold of 0.5), which compared to the baseline of 29 percent (the current percentage) means that this model is roughly 30% percent better at identify projects that might fail than a random test.

If DonorsChoose is interested in a model to identify 5% of posted projects that are at highest risk of not getting fully funded to intervene with, you would want the model that had the best precision at 5%. This would be logistic regression (parameter C=1), which has a precision of 53.02 percent at this level. Targeting this group with resources would allow for significant bang

---

[1] Bruce J. Biddle and David C. Berliner, "A Research Synthesis / Unequal School Funding in the United States" http://www.ascd.org/publications/educational-leadership/may02/vol59/num08/Unequal-School-Funding-in-the-United-States.aspx

[2] Jill Barshay, "While the rest of the world invests more in education, the U.S. spends less", https://hechingerreport.org/rest-world-invests-education-u-s-spends-less/

for your buck as slightly over half of the people you targeted would actually have their project fail (according this model).

Recommendations for DonorsChoose

DonorsChoose could reach out to those who the model identified as likely not having their project be funded within 60 days and offer them support in advertising their project.
The model indicated certain features were related to a higher likelihood of being chosen. For example, having a project be eligible for a match was related with higher funding. Donor's Choose could recommend teachers more proactively try and seek these matching funds to boost overall funding.

Top 10 evaluation results, sorted by precision at 5 percent:

	model	parameters	f1_score	precision_at_0.5	recall_at_0.5
23	logistic_regression	{'C': 1}	0.534307	0.530231	0.107869
21	logistic_regression	{'C': 0.1}	0.531499	0.530106	0.099254
22	logistic_regression	{'C': 10}	0.534671	0.529746	0.109202
20	logistic_regression	{'C': 0.01}	0.522528	0.523623	0.072110
3	boosting	{'n_estimators': 100}	0.529340	0.522228	0.093926
5	boosting	{'n_estimators': 50}	0.527909	0.520514	0.089850
4	boosting	{'n_estimators': 30}	0.524234	0.503572	0.081067
10	decision_tree	{'criterion': 'entropy', 'max_depth': 8}	0.526144	0.474059	0.097283
19	logistic_regression	{'C': 0.001}	0.503724	0.473304	0.012972
17	knn	{'n_neighbors': 25}	0.530774	0.469976	0.114557