Short Report

Si Young Byun

1. Problem / Goal

DonorsChoose.org is an online charity website that allows to help students in need through donation. The goal is to help identifying 5% of posted projects to intervene with by running a machine-learning analysis on the dataset projects on *DonorsChoose.org* to predict which projects are fully funded within 60 days of posting.

2. Machine learning Models

Seven classifiers (i.e. Logistic Regression, K-Nearest Neighbor, Decision Trees, SVM, Random Forests, Gradient Boosting, and Bagging) have been used to identify the best model for this goal. We will be using several evaluation metrics to evaluate these classifier models.

3. Results

	Date	Model	Parameters	Recall_at_50%
12	2013-01-01	GB	Default	0.588422
19	2013-07-01	GB	Default	0.579047
17	2013-07-01	SVM	Default	0.572686
10	2013-01-01	SVM	Default	0.566646
5	2012-07-01	GB	Default	0.56301

Because we care about helping projects based on the prediction of whether the projects will fully funded within 60 days of posting, it is important that we are not missing any projects that require help. Therefore, we propose that they should choose **Gradient Boosting** to deploy as it has the best recall scores across the three temporally split test data since recall looks at what proportion of actual positives was identified correctly by the model.