
CS 4780 Final Project Proposal

1. Team

This project will be completed by a team of four students.

2. Motivation

DonorsChoose is a crowd funding website that helps public school teachers request and receive funding. 70 percent of campaigns on DonorsChoose are successfully funded. While this is a healthy margin, it could be vastly improved. It would be valuable for teachers to learn what factors may affect the success of their campaign.

Extensive research about Kickstarter and other similar crowd-funding sources has already been done to investigate causes of successful commercial campaigns. It would be interesting to compare and contrast factors that influence the successful of a crowd-funding campaign in the commercial realm versus in the philanthropic realm.

3. Problem Statement

The main goal of this project is to determine what factors have the greatest influence on if a project will be fully funded. In particular, we are interested in investigating whether characteristics of a project such as the location of the school, the poverty level, the grade level, or area of study (such as english versus chemistry) affect the likelihood of funding.

We are also interested in looking at how characteristics of already pledged donations affect likelihood of future donations, and thus success of projects in a time-series framework.

DonorChoose enables various promotions such as having a corporation match donations. We would like to investigate if these promotions affect the number or size of donation as well as if they affect the likelihood of a project being funded.

The dataset also includes project description essays, written by the creating teachers. We'd like to use some existing software to extract important keywords and add these as feature vectors for our linear classifier.

4. Approach

We are trying to find a linear classifier that describes what a successful project entails, highlight which "features" are most important, and try to understand the effect of "momentum" in terms of donations.

We plan to use existing software to find a linear classifier (like SVM light) and will experiment with different scaling of our data as well as omitting certain fields to achieve the best outcome.

As for identifying the most important features we will look at the weighting of the feature vectors in the optimal hyperplane as well as run some model selection tests where we omit certain features and observe the change in accuracy on our test set.

In order to evaluate the effect of "momentum" we will try adding this data in during different parts of the funding process to see if SVM light can find a connection as well as model the correlation between amount and date of funding and the ultimate success of the project.

5. Resources

The full dataset for this project is provided publicly by Kaggle. It is available here: <https://www.kaggle.com/c/kdd-cup-2014-predicting-excitement-at-donors-choose/data>.

We will use existing linear classification software provided publicly online, starting with SVM light and expanding to other software as necessary.

File reading and other custom code written for this project is written in Scala, a publicly accessible and OS-agnostic language.

6. Schedule

- 22nd October: This proposal submitted
- 25th October: Process data, begin classification
- 2nd November: Compile results, begin comparison of different results
- 9th November: Compile models, begin compari-

110	son of different models	165
111		166
112	• 11th November: Submit progress report.	167
113		168
114	• 16th November: Being writing poster and report.	169
115		170
116	• 4th December: Poster presentation. Submit poster.	171
117		172
118	• 5th December: Peer reviews for posters due.	173
119		174
120	• 10th December: Final project report (and code) due.	175
121		176
122	• 15th December: Peer reviews for final project reports due.	177
123		178
124		179
125	• 16th December: Author Feedback on reviews for final project reports due.	180
126		181
127		182
128		183
129		184
130		185
131		186
132		187
133		188
134		189
135		190
136		191
137		192
138		193
139		194
140		195
141		196
142		197
143		198
144		199
145		200
146		201
147		202
148		203
149		204
150		205
151		206
152		207
153		208
154		209
155		210
156		211
157		212
158		213
159		214
160		215
161		216
162		217
163		218
164		219