Stacey Camp

Final Capstone Proposal

1/12/2018

**Title**: Predicting Successful Funding of GlobalGiving Projects

**Problem**: GlobalGiving is an online platform for crowdfunding charitable projects. Non-profit organizations, companies and individuals can use the GlobalGiving site to make requests for donations toward completing specific projects. The status for inactive projects is either “retired” or “funded.” Using this status as a target variable, the goal of this project is to develop a set of features from the project data that is predictive of whether or not a project will be fully funded.

**Value**: There are a lot of big problems in the world. There are also a lot of people out there trying to solve them. GlobalGiving (and similar crowdfunding platforms like GoFundMe, Causes and Fundly) help people and organizations raise awareness and funds for their particular cause or interest. In order for GlobalGiving to compete effectively with other charitable crowdfunding sites and to generally be more successful in their mission to make the world a better place, they need to maximize the likelihood that their hosted projects will be funded. By identifying the characteristics of a successful project, GlobalGiving could strategically feature projects with a high likelihood of success to promote faster turnover in projects. This strategy could secondarily increase visibility to other projects that would no longer be competing against more popular projects. Given that there are currently over 4000 active projects on the GlobalGiving site, this output could also be used to counsel project owners on how to develop a competitive funding request. Finally, a model developed using GlobalGiving data could potentially be used as a predictor for other charitable crowdfunding sites or even sites raising money for crowdfunding business startups or product development.

**Data Source**: GlobalGiving provides an API with access to data for both active and inactive projects. I intend to use the data obtained using the “Get All Projects” and “Get Featured Projects” methods to gather information for all of the projects that have been hosted on their site and incorporate information on whether that project has ever been featured on website main page.

**Techniques**: The majority of time spent on this project will likely be in feature engineering. I’ll initially use clustering, descriptive statistics and visualization techniques to identify patterns that may lead to useful features. NLP techniques such as tf-idf and bag-of-words may be used to derive features from the text based fields such as the project summary, donation descriptions and organization missions. Various supervised learning models will be tested for their effectiveness in predicting outcome. Finally, time allowing, I’m interested in using survival analysis techniques to predict time to funding.

**Challenges**: Successfully engineering predictive features for this dataset using mostly text and numeric data could be challenging. The content of images likely has a huge impact on getting potential donor’s attention and driving them to provide funding. The only data output that will represent this data is the number of images contained in the projects image gallery. Furthermore, the API doesn’t provide data on donors so information related to monthly versus one-time donations while available on the website is not provided as output from the API. This is another source of data that is likely highly relevant to the outcome of the project but unavailable.