Project Proposal

Prepared for: Capstone project 2(Data Science Career Track)

Prepared by: Vishal Kumar

March 7, 2018

Proposal number: 1

PROJECT PROPOSAL

Introduction

DonorsChoose.org is a United States–based 501(c)(3) nonprofit organization that allows individuals to donate directly to public school classroom projects. Founded in 2000 by former public school teacher Charles Best, DonorsChoose.org was among the first civic crowdfunding platforms of its kind. The organization has been given Charity Navigator's highest rating every year since 2005. In January 2018, they announced that 1 million projects had been funded. In 77% of public schools in the United States, at least one project has been requested on DonorsChoose.org. Schools from wealthy areas are more likely to make technology requests, while schools from less affluent areas are more likely to request basic supplies. It's been noted that repeat donors on DonorsChoose typically donate to projects they have no prior relationship with, and most often fund projects serving financially challenged students.

DonorsChoose.org receives hundreds of thousands of project proposals each year for classroom projects in need of funding. Right now, a large number of volunteers is needed to manually screen each submission before it's approved to be posted on the DonorsChoose.org website.

Next year, DonorsChoose.org expects to receive close to 500,000 project proposals. As a result, there are three main problems they need to solve:

1. How to scale current manual processes and resources to screen 500,000 projects so that they can be posted as quickly and as efficiently as possible

2. How to increase the consistency of project vetting across different volunteers to improve the experience for teachers

3. How to focus volunteer time on the applications that need the most assistance

Goals

The goal is to build a classification prediction model that predicts whether or not a DonorsChoose.org project proposal submitted by a teacher will be approved, using the text of project descriptions as well as additional metadata about the project, teacher, and school. Donors-

Choose.org can then use this information to identify projects most likely to need further review before approval.

With an algorithm to pre-screen applications, DonorsChoose.org can auto-approve some applications quickly so that volunteers can spend their time on more nuanced and detailed project vetting processes, including doing more to help teachers develop projects that qualify for specific funding opportunities.

This model can be used by any fund raiser companies as a filter to select certain qualifying projects to be posted on the internet. For example, online fund raiser companies would get hundreds of request everyday and they must analyze projects importance and need, before they can be posted on the website, this model will help them reduce the time spent on analyzing these project manually.

I believe this model can also be used by public school teachers to prepare better proposal and submit it to the DonorsChoose.org.

About Data

The dataset contains information from teachers' project applications to DonorsChoose.org including teacher attributes, school attributes, and the project proposals including application essays.
https://www.kaggle.com/c/donorschoose-application-screening/data

Project Outline

I will use different machine learning classification algorithms and compare the models for best result for the above problem. Since the problem has depended variable of text data, Natural Language Processing(NLP) will play very important part in building the model.

Project Deliverables

Jupyter notebook: which will have all the codes, plots, links(if necessary), associated documents, project report and Slide deck.