

Policy Memo – Donors Choice Projects

Selecting the projects that are least likely to be fully funded

Donors Choose is an organization through which teachers and school administrators can crowd source funding for classroom endeavors. Their requests are varied, and can include anything from a need for classroom equipment, to funding for special trips, or the means with which to start a new project. While the majority of projects are funded, about 30% are not. The nature of crowd sourcing makes it difficult to know what attracts potential donors eyes (and wallets), or any distinguishing factors that prevent certain projects from being fully funded. Knowing this information would help donors choose employees and requesters to frame projects in a way that increases their projects' likelihood of being chosen. Alternatively, if there is a reason that certain projects are not chosen, Donors Choose could elect to market those different or to an audience that is more likely to fund them.

This analysis was done on donors choose projects from 2011 – 2014. Data was received both about the projects that were initially pitched and their defining characteristics, and the outcome of the requests – whether or not they were fully funded, and some of the factors that went into that decision making process.

The data on the projects contained identifying information on the projects, details about the school, such as whether it was a magnet school or a charter school, and more information about the project, such as the grades it served and the subject area it was for. Data was also provided about the final outcome, noting whether or not the project was fully funded, outstanding factors about the donors, teacher references, and curiously enough, a note about whether or not the project was considered “exciting”.

Initial analysis showed the states where the most number of projects are requested are California, New York, Texas, Florida, and North Carolina. This is not surprising as these are some of the larger states, and the ones who often have problems with public school funding, especially in urban areas. However, the most projects are funded in the first four states, and then in Illinois. The projects (as can be seen in the bar graphs in the analysis) appear consistently distributed across grade levels, poverty levels, and although there is some more variation here,

subject area. This suggests that people are not more or less likely to donate to projects based on the factors.

The next step of the data analysis was to compare the projects that were funded with those that were not. While the average amount requested for a project was \$531, the average amount requested for the projects that were fully funded was \$482. Since Donors Choose relies on a number of individual donations, it seems reasonable that a more expensive project may be more difficult to fund. An unusual factor from the 'outcomes' table was 'is_exciting', which was an assessment of whether or not a donor found a project exciting after a conversation about it. This did not prove to be a strong indicator, as only 13% of the funded projects were deemed "exciting". However, none of the not funded projects were deemed exciting.

Another factor that was examined was whether or not there was a teacher referred donor. 23% of unfunded projects had a teacher referred donor while 39% of the funded projects had at least one. Referral from a teacher seems to be a contributor to making sure a project is funded. This is likely a signal for quality and increases the visibility of the project. It may also be a proxy for other factors as these projects may have more support in the school of a more qualified teacher supporting it.

Following the initial analysis a series of classification models were created. The models aimed to predict which projects were more likely not to be funded based on the characteristics of projects that had been funded in the past.

Prior to the classification the data provided was further processed. The mean value was imputed for 'students_reached' rows where the value there was not present. Three of the other numerical categories were divided into price ranges and turned into binary variables.

A series of machine learning techniques were applied in three shifted time frames. In each time frame the classifier was trained on one year of data, and then tested on the following six months. This would allow the final model to be optimized to predict projects in the next 6 months.

A following consideration was which would be more important – precision or recall. In this case, it seems more important to be able to accurately pinpoint projects that were definitely fully funded, even if every funded project is not captured by the model. Therefore more attention was paid to the precision metric. The models also performed better overall on precision compared to recall.

The models can be compared using the AUC-ROC score, an overall measurement of the evaluation metrics mentioned above. None of the models had a high rate, but the Logistic Regression performed best, at about 0.65, for all the time periods. This suggests that the Logistic Regression model should be used to make predictions for future projects.

However the overall low performance of the models suggests that more should be done in order to create a more effective model. This could include varying the parameters further within the classifiers, or considering more factors. Since a high percentage (70%) of the projects are already fully funded, it would be important to examine whether the current donor base has the funds for every single project. Further, project funding may depend on more subjective factors such as how well the project is presented or written up that would not be captured by the model. Following this analysis the recommendation is both to look for additional factors that could be added to the model for analysis, but also to consider the effect of 'softer' or less technical factors on project funding.