Project Proposal

Karissa Dunkerley, Lucia Fang, Sen Feng

Project Name: Predicting DonorsChoose Project Funding Success Based on Poverty Levels

Project Selection

We selected the DonorsChoose Predicting Excitement project instead of the Tanzanian Ministry of Water Pump it up project mainly due to our interest in the topic of education and its potential impact on educational equity. While both projects could have a substantial impact, the DonorsChoose project resonates more with us as students deeply invested in education. We value quality education. Thus, knowing that the educational experience of students in districts that cannot financially afford to provide for all of their classrooms' needs may suffer due to a lack of resources motivated us to decide on DonorsChoose for our project. By leveraging machine learning techniques, we believe we can substantially improve the effectiveness of this crowdfunding platform, ultimately benefiting students in high-need areas.

Background and goals

Teachers, particularly in low-income areas, often pay out-of-pocket for classroom supplies because the schools lack the necessary resources to provide for a quality education, leading to a reliance on external funding platforms like DonorsChoose.org. On average, teachers spend around $450 per year on school supplies (Tahmincioglu, 2020). DonorsChoose.org allows teachers to seek donations for their projects. On this website, teachers can post requests for classroom resources, along with a brief description of what they need and what it will be used for to entice donors to contribute. Then, they can share their request with friends and family, as well as the community of donors cultivated by DonorsChoose.org. Projects have four months to be funded on the site and if not funded, donors can redirect their contributions toward that teacher's next project. Thousands of teachers post projects on this site. However, many projects fail to reach their funding goals, depriving students of educational resources.

In fact, approximately 30% of DonorsChoose resource requests are not funded within the allotted four months, some of which come from high poverty areas. This means that many classrooms in high-need areas go without essential educational materials, despite the filters and analytics DonorsChoose has in place to promote the requests to donors. Currently, donors can find projects to support based on if they are coming from historically underfunded schools due to economic and racial inequity, if the requests are classroom essentials, metrics to maximize their impact, the amount needed, the topic of the request, the grade level, the supplies needed, and even if the request is an experience. Another method DonorsChoose.org uses to prioritize their funding is identifying high quality/exciting projects to promote to their donors on the home page. However, they define "exciting" projects based on donor connection type (e.g., teacher referred donor) and donation quality. Because teachers from higher need areas might not know that many wealthy donors, this could be an inherent bias that propagates through the prediction stages. As shown in Appendix I.4, there are proportionally fewer fully funded projects that were considered "exciting" in high and highest poverty areas when compared to low and moderate poverty areas. This issue exacerbates educational inequality as students in these areas are already at a disadvantage compared to their peers in wealthier regions.

Addressing this issue is crucial because increased funding for educational resources has been linked to improved student outcomes, including higher graduation rates (Warren, 2009; Houck, 2010). Teachers from schools in high-poverty areas could request for very different resources than their counterparts. When these projects go unfunded, students miss out on crucial learning opportunities, potentially impacting their educational outcomes and future prospects. Ensuring that projects from high-poverty areas are funded is vital for promoting educational equity and providing all students with a fair chance at success.

Therefore, at a high level, we hope to increase funding for underfunded projects. More specifically, we would like to increase the number of projects reaching their funding goals from low income areas, due to the greater impact funding those projects can have. This targeted approach will ensure that resources are allocated more efficiently and equitably, helping to bridge the gap between high-poverty and wealthier schools. The impact will be seen in better-equipped classrooms, improved learning experiences, and, ultimately, a reduction in educational inequality. Furthermore, teachers and students in high-poverty areas directly benefit from increased funding, leading to improved educational resources and outcomes. Additionally, solving this problem will help DonorsChoose enhance the platform's ability to fulfill its mission of supporting education by ensuring a more equitable distribution of resources. Moreover, donors will gain insights into where their contributions can make the most significant impact, particularly in supporting underfunded schools. Educational Policymakers can use data-driven insights to inform funding decisions and create more effective policies aimed at reducing educational disparities.

Policy Goals

We focus on equity, efficiency, and effectiveness in increasing the amount of projects from low socioeconomic areas in being funded. For equity, we want to forgo the ‘exciting’ labels that could unintentionally favor wealthier schools, often influenced by a teacher’s network. By prioritizing equity, we aim to create a fairer platform where projects from high-poverty areas have a better chance of being funded. For effectiveness, we will elucidate the resources that are most prominent in fully-funded projects in each poverty level to make the most use out of DonorChoose.org project labeling. For efficiency, we will optimize the allocation of donor resources to maximize educational impact. While focusing on equity by prioritizing high-poverty areas, we must also consider the efficient use of resources. Our approach aims to maximize impact by identifying projects that, without intervention, are unlikely to be funded, ensuring a more effective allocation of donations. To address potential biases, our model must be designed to avoid any biases that could unintentionally disadvantage certain projects or schools. We will incorporate fairness metrics to ensure that the solution promotes equitable outcomes.

Ethical Considerations

There are five main stakeholders in our project: teachers, students, donors, DonorsChoose.org, and educational policymakers. First of all, teachers need to know if their project is being highlighted for funding by DonorsChoose.org as a result of this ML project and they need to be provided the insights to make their project descriptions more attractive to donors. The teachers’ students would then experience the benefit of their teachers’ projects being highlighted for funding. Furthermore, the donors should be able to see what projects are unlikely to be funded within the four months so that they can donate to these projects. The donors should also be able to see the methodology behind highlighting the projects somewhere on the website for transparency. Moreover, DonorsChoose.org should understand the whole of the project, from the goal to the model to the results to how to integrate the model into their system. This will allow them to use the model to identify projects that are likely to struggle with funding and specifically promote those on the website. To facilitate their understanding, we will try to keep our analysis less complex so it can be understood more easily by non technical people and more easily integrated into their existing system. Lastly, educational policymakers should be able to see the results of our project because areas with projects that struggle to be funded might correlate with specific attributes of an area.

Thus, these areas could be in need of more government funding and grants. We also should give these policymakers access to where our data came from so they can be transparent about it with their constituents. Finally, to preserve privacy, we will ensure that no personally identifiable information about students or teachers is used in our models. All data will be anonymized and aggregated to protect individual privacy.

Actions

One action is for DonorsChoose to add a section to the website's home page to promote the most likely projects not to be fully funded from high poverty areas. This could be accomplished by their web developers adding a section called something like "Projects of the Week" to the home page, where they have these projects listed on the home page for donors to see. The listed projects will be updated once a week by DonorsChoose.org, as they are the ones who need to promote the projects on the website, and that is how often the list of projects will be generated by the ML project. Another action for DonorsChoose web developers is to add a filter to the search for projects page on the website that could address the projects from high poverty areas unlikely to be fully funded. The filter could be under the MAXIMIZE YOUR IMPACT section of filters. Moreover, the projects that this filter selects should be updated weekly as a new list of projects unlikely to be funded is generated from the ML model.

Additionally, DonorsChoose can periodically recommend these projects from low income areas unlikely to reach their funding goals when a donor elects to have the website choose a project for them to support. To do so, the DonorsChoose software developers would have to incorporate these projects by priority (the higher the risk of not being funded, the higher the likelihood of being recommended) into their algorithm that chooses which project to recommend to the donor. Because the list of projects is created weekly from the ML model, the projects available to be recommended should be updated weekly as well. Currently, DonorsChoose aims to promote the funding of projects from areas with lower socioeconomic status by 1) having a section on the home page for Equity Focus School Projects and 2) having filters for historically underfunded schools due to economic and racial inequity, never before funded teachers, more than half of students from low-income households, and projects with no donations.

Data Analysis

We need to explore variables that may be correlated with whether or not projects from high/highest poverty areas achieve full funding within the four month time constraint. The DonorsChoose data is made up of five main csv files: donations, essays, outcomes, projects, and resources. The donations file has information (donation amount, donation message, donor type, etc.) about all of the donations and donors for the projects, one row per donation. The essays file contains the information about the teachers' request proposals for their projects to be placed on the website for funding, one row per request proposal. The outcomes file is the outcomes for the projects, whether they are funded, considered excited, etc., with one row per project. The projects file contains all the information about the projects, what school they are from, what is the poverty level of the school, the subject of the project, the type of resources requested, etc., one row per project. The resources file contains the information for all of the resources requested for the projects, like what the vendor is, the project resource type, the item name, the item price, etc., one row per resource.

We first looked at some feature correlations. From the heatmap in Appendix I.I, we can ascertain several key insights, like the identification key predictors and where we might need to reduce multicollinearity, so we can start to see which variables might be useful in predicting whether a project gets funded or not. For instance, it shows a correlation between the poverty level encoded (1 corresponds to low poverty, 4 corresponds to the highest poverty level) and other numerical values. As shown, there is a strong positive correlation between donation_total (the amount the donor gave toward the project) and donation_optional_support (the optional support funding the donor gave toward the project) which means that as donation_optional_support increases, the donation_total will tend to increase as well. Therefore, to address this issue, we might need to combine both variables into one variable or only use one variable in our machine learning model. By pinpointing the features that most significantly impact a project's likelihood of success, we can more accurately identify projects at risk of not being funded.

Then, we added a variable called funded_in_4_mths, which is a boolean variable that describes whether or not the project was funded within the 4 month required time frame. We also explored the counts of unfunded projects from high/highest poverty areas for the project’s primary subject (the school subject the project relates to) and resource type (the type of resource of the requested project). In the bar charts in Appendix I.2 and I.3, we can see that many of the unfunded projects had literacy as their primary subject and had supplies and technology as their resource types. On their own, we cannot draw significant conclusions from these graphs, but we can explore how the counts for each subject and resource_type from the unfunded projects compare to those of the funded projects to see if there are any correlations.

Additionally, we looked at the is_exciting variable (which is a boolean that expresses whether a project is exciting or not in business terms) to see if there were any differences in which projects were labeled excited between the poverty levels. The bar chart in Appendix I.4 shows that in higher poverty areas, the proportion of fully funded projects considered "exciting" is a little bit lower compared to lower poverty areas. For instance, in high poverty areas, about 16% of fully funded projects are considered exciting, whereas 84% are not. This trend continues in the highest poverty areas, where about 15% of fully funded projects are exciting. In contrast, lower poverty areas have a higher proportion of exciting projects, with about 17% exciting and 83% non-exciting. This indicates that there could be a slight difference in the visibility and attractiveness of projects from high and highest poverty areas compared to those from lower poverty areas.

Moreover, we looked at the donation_message variable (messages that donors submitted along with their donations) from the donations file to see if there was a correlation in the words used for projects that were funded. The analysis of donation messages shows that donors often talk about supporting education and helping students, using words like "gave," "education," "support," and "kids." This tells us that donors are moved by the idea of directly impacting students' learning, and this could be valuable in predicting whether or not a project is likely to be funded. From this analysis, the top words for funded projects are: gave, project, education, support, kids, want, students, strong, and believer.

Some other variables that may factor into whether or not a project from a low income area is fully funded may be at_least_1_teacher_referred_donor (did the project have at least one teacher-acquired donor), teacher_referred_count (the number of donors that were teacher referred), school_metro (whether the school is rural, urban, or suburban), students_reached (number of students impacted by a project if funded), is_teacher_acct (donor is also a teacher), title (title of the project), and short_description (description of a project).

One outcome variable we might use is funding_likelihood, which would be a likelihood score/probability from 0 to 1 of a project’s likelihood of being funded.

Analytical Formulation and Metrics

Every week, for all active DonorsChoose.org projects posted that have not yet reached their funding goal, we can identify the top 10 projects that are both at high risk of not reaching their funding goal within 4 months and are from schools with high socioeconomic need, in order to prioritize these projects for additional promotion and increase the number of projects being fully funded from low income areas, whose schools and students may not be able to fund the projects themselves.

For metrics, we will focus on recall (or sensitivity) as we need to measure how well the model identifies true positives regardless of the number of false positives. In this context, the cost of incorrectly identifying a project as high-need (false positive) is relatively low compared to missing out on funding critical projects (false negatives). A false positive could mean a project that might not be in as needed is promoted, but if this results in more funding overall, the impact could still be positive. On the other hand, a false negative means a genuinely disadvantaged project misses out on funding, which could directly harm students. In addition, we understand the need to balance with precision to ensure we're not promoting too many projects that don't actually need the extra boost. To measure the success of our approach, we will use metrics such as the funding rate for projects with a high/highest-poverty level, which is the percentage of projects that reach their funding goal within four months. We could also use an equity index, which would be the measure of funding distribution across different poverty levels, which will help us make sure that while we are promoting projects from areas with the most socioeconomic need, it will not negatively affect the funding given by donors to projects in areas with low poverty. By tracking these metrics, we can assess whether the machine learning model is effectively promoting equity by increasing the visibility and funding likelihood of projects in high-poverty areas.

Additional Considerations

Caveats

Our analysis is limited to the data available through DonorsChoose.org. We may lack insight into external factors affecting donation patterns such as donor preferences. Additionally, our findings may not generalize to other educational funding contexts or platforms and we may be limited in the complexity of the model we can create, due to the time constraints of the semester.

Policy Recommendations

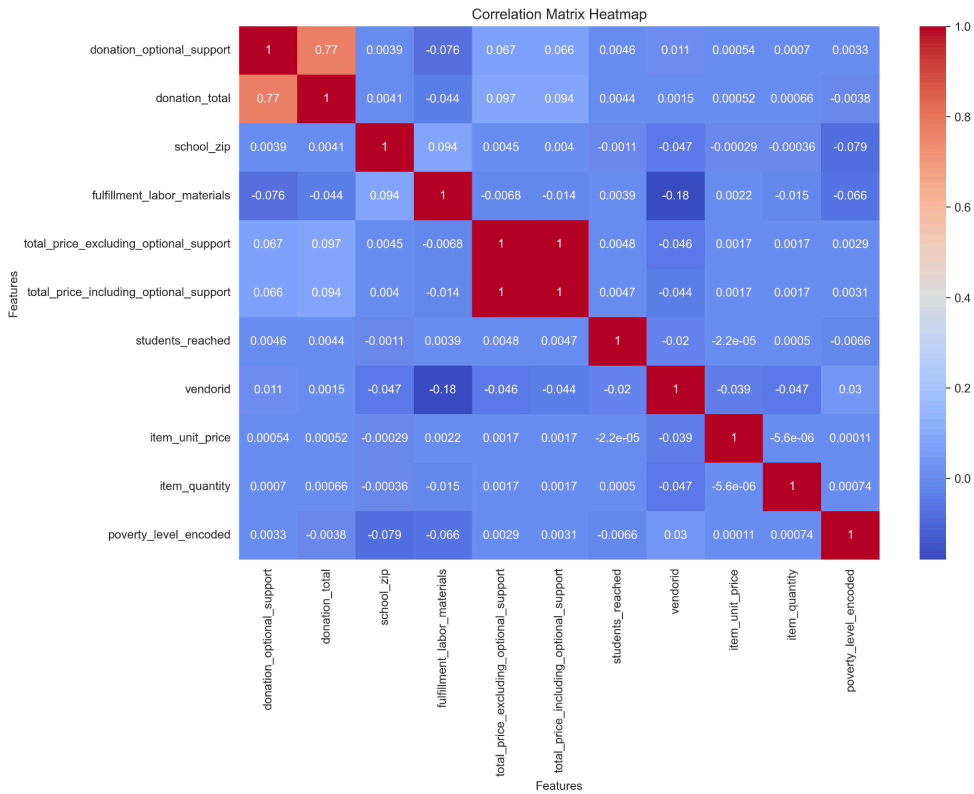
Based on this analysis, we recommend that policymakers at DonorsChoose implement a machine learning model tailored for each poverty level on DonorsChoose.org to prioritize projects more likely to be funded. By pushing these projects to the top of the listings, we can accelerate their funding process, ensuring that resources reach classrooms in need more quickly. Another recommendation is to adjust the donor recommendation algorithm to periodically promote projects from low-income areas that are unlikely to reach their funding goals, which will help ensure that high-need projects receive the attention and support they deserve, even when donor-teacher connections or biases might otherwise hide them.

To validate the impact of these recommendations, we will track several key metrics over time, including the increase in funding rates for projects in high-poverty areas, the reduction in time-to-funding for prioritized projects, and the number of at-risk projects that become fully funded due to the new recommendation system. A/B testing can be used to compare the current system with the proposed changes to ensure that the new model is effectively increasing equitable access to resources. Additionally, we will gather feedback from teachers and donors on the new system to ensure it meets their needs and fosters a more inclusive environment for project funding.

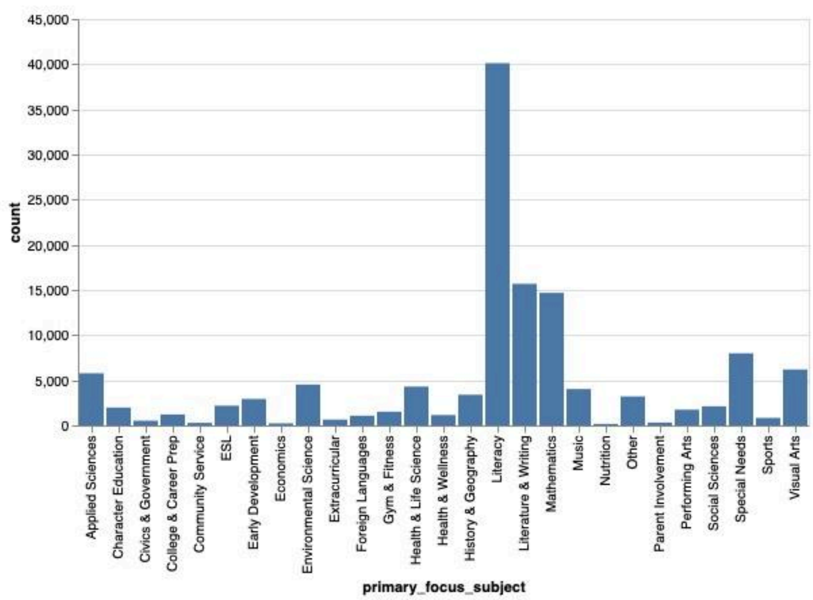
For educational policymakers, the fact that there are so many funding requests from high poverty areas suggest that the schools are not being adequately funded. This information could be used to help policymakers discern which areas need more funding and work to implement policies to do so. We will validate whether our proposal will have the desired impact if there is a substantial increase in projects from high poverty areas being funded, and also at a very high level if there are new policies implemented to address the inequity in public school funding.

Appendix I: Graphs

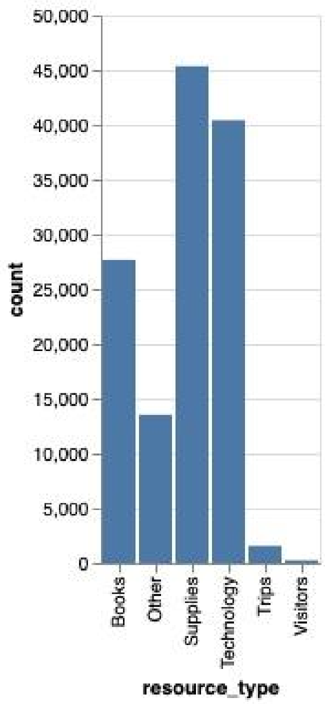
I.I Heatmap



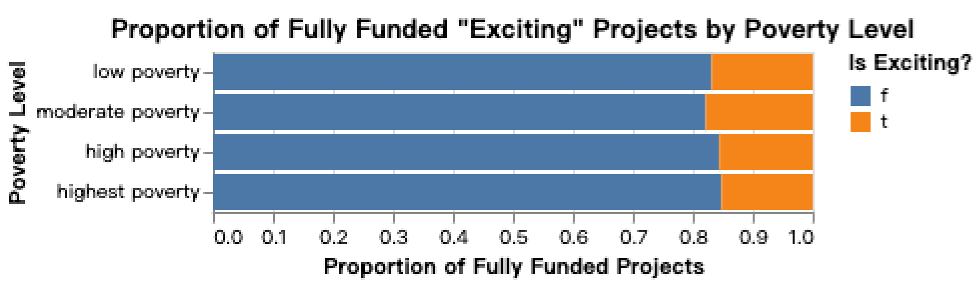
I.2 Bar Chart of primary_focus_subject



I.3: Bar Chart of resource_type



I.4: Proportion of Fully Funded "Exciting" Projects by Poverty Level



Appendix II: References

Houck, E. A., & Kurtz, A. (2010). Resource Distribution and Graduation Rates in SREB States: An Overview. *Peabody Journal of Education*, *85*(1), 32–48. https://doi.org/10.1080/01619560903523698

KDD Cup 2014 - Predicting Excitement at DonorsChoose.org | Kaggle. (2014). Kaggle.com. https://www.kaggle.com/competitions/kdd-cup-2014-predicting-excitement-at-donors-choose/overview

Tahmincioglu, E. (2020, December 3). *Teachers pay out-of-pocket to keep their classrooms clean of COVID-19 : Teachers already spend on average $450 a year on school supplies*. Policycommons.net; Economic Policy Institute. https://policycommons.net/artifacts/1408189/teachers-pay-out-of-pocket-to-keep-their-classrooms-clean-of-covid-19/2022454/

Warren, J. R., & Halpern-Manners, A. (2009). Measuring High School Graduation Rates at the State Level. *Sociological Methods & Research*, *38*(1), 3–37. https://doi.org/10.1177/0049124109339374