Demystifying Social Entrepreneurship : An NLP Based Approach to Finding a Social Good Fellow

Aditya Garg[1], Kush Varshney[2], Alexandra Olteanu[2], Richard B Segal[2], Dmitriy Katz-Rogozhnikov[2], Keerthana Kumar[2], Joana Maria[2], Liza Mueller[3], Ben Beers[3]

1- IBM Social Good Fellow, Columbia University, 2 - IBM Research. 3 - Echoing Green

ABSTRACT

Social entrepreneurs who start organizations to tackle social, cultural, or environmental issues, are often lost early in their journey due to lack of visibility in the space of sustainable development. Echoing Green—a funder of social enterprises—annually receives around 3000 applications from such social entrepreneurs, and on average selects just over 1% of them for a two-year Fellowship that includes both financial and leadership development support. The overall collection of applications is a rich data source for gaining insights about what makes applicants, their idea and their organization successful in achieving social impact; but also, to draw insights about topical trends and their evolution over time. To study this, we use machine learning and natural language processing techniques. Specifically, in addition to available responses to categorical questions regarding applicants e.g., educational background, their organization funding, or online presence; we also explore and extract a variety of lexical, syntactic, semantic, and personality-related features from detailed textual responses about the applicant, their idea, their solution, and their organization. We then use these features to build a classification model that is able to separate successful applications from the rest of the application pool, using a master dataset that combines six years of applications. We also identify features that are more commonly found in successful applications.

1. INTRODUCTION

Social Entrepreneurship has witnessed growth throughout the world in the last years,[1] as many entrepreneurs increasingly see social good as a credible path to drive both profit and system change through innovation. These entrepreneurs are typically backed by accelerators and venture funds, who provide resources, mentorship, and guidance in the early stages of these social initiatives. One such venture is **Echoing Green** (*www.echoinggreen.org*), a social innovation fund that invests in emerging leaders and their social initiatives. Through its fellowship program, **Echoing Green** fosters social good by identifying leaders in the social entrepreneurship domain—providing them a seed fund, a structured leadership development program, and a community of peers. To shortlist this group of fellows, **Echoing Green** manually vets over 3000 applications through its long and rigorous 3-phase application evaluation process; a process that is labor intensive and highly time consuming, taking about 6 months to complete.

Echoing Green believes in investing in and supporting the right people relative to the right ideas and ability to execute, rather than supporting specific business plans. Therefore, finding the right applicant is more valuable than finding the right idea. Fellowship applications contain comprehensive information about both the applicants and their initiatives. This rich textual data

[1] Global Entrepreneurship Monitor's (GEM) Social Entrepreneurship Report released at the end of May 2016

provides us a unique opportunity to analyze the pool of applicants to gain insights about what goes into making a successful application—awarded the Echoing Green Fellowship.

Thus, one of our goals is to identify common traits found in successful applications. To this end, we use machine learning and natural language processing techniques to extract as much information as possible from the application data through feature engineering and exploration, features that aim to characterize the ideal traits or features that are requisite to be a successful application. The results of this study can support application reviewers at **Echoing Green** who spend significant time evaluating the applications.

To make the application evaluation process bias free, fair, and equitable for all, **Echoing Green** also dedicates significant resources on evaluator training. Yet, subtle, unknown biases may still seep in and affect the evaluation process; and, thus, identifying and addressing them remains a sustained effort over time. Therefore, another goal is to check for any unintentional biases in the evaluation process, and highlight their impact on the success of an application.

Finally, one of the most laborious task in the application evaluation process is shortlisting applications in the first phase of the process, where about 80% of applications are rejected after they are reviewed by multiple evaluators. To this end, we explore if we can train a classifier that accurately predicts the probability of an application to be accepted in their initial evaluation phase. **Echoing Green** could use such an automated process to highlight specific applications for focused review, or to help evaluators while classifying applications.

2. METHODOLOGICAL FRAMEWORK

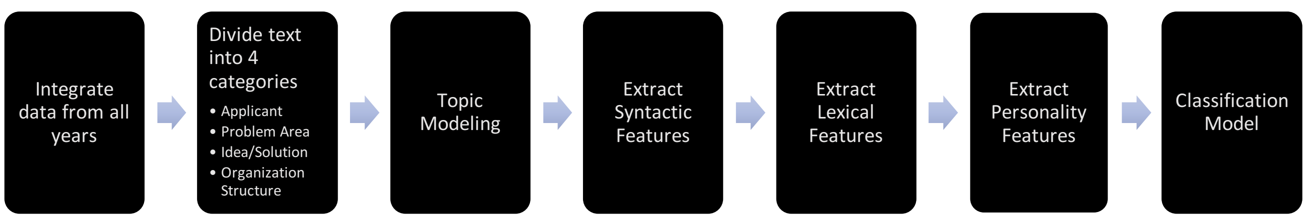


Fig. 1: Overview of methodological framework

In this section, we briefly overview our data collection covering 6 years of applications to Echoing Green, our extensive list of features, and show how we can use these features to automatically classify the applications. We also report on our classification model performance.

2.1 Data

Our data consists of the filled fellowship application forms from 2012 to 2017. All applications are labelled with their final qualification status which could be take the following possible values: **Rejected** – rejected in Phase 1; **Semi-Finalists** – qualified through Phase 1; **Finalists** – qualified through Phase 2**; Fellows** – selected for the Fellowship.

Year	Total Applicants	Semi-Finalists	Finalists
2012	3500	418	28*
2013	2863	444	53
2014	2726	444	85
2015	3165	433	79
2016	2077	425	58
2017	2879	436	56

Distribution of applicants across years (*Number of Fellows, since Finalists information is not available in 2012)

In the application form, Echoing Green asks the applicants to describe their background and experience, the problem area they are addressing, the solution they are proposing and their organizational structure in the form of free text questions. Besides these free text questions, the form also contains categorical input about the applicants' demographics and their organization.

2.2 Feature Exploration

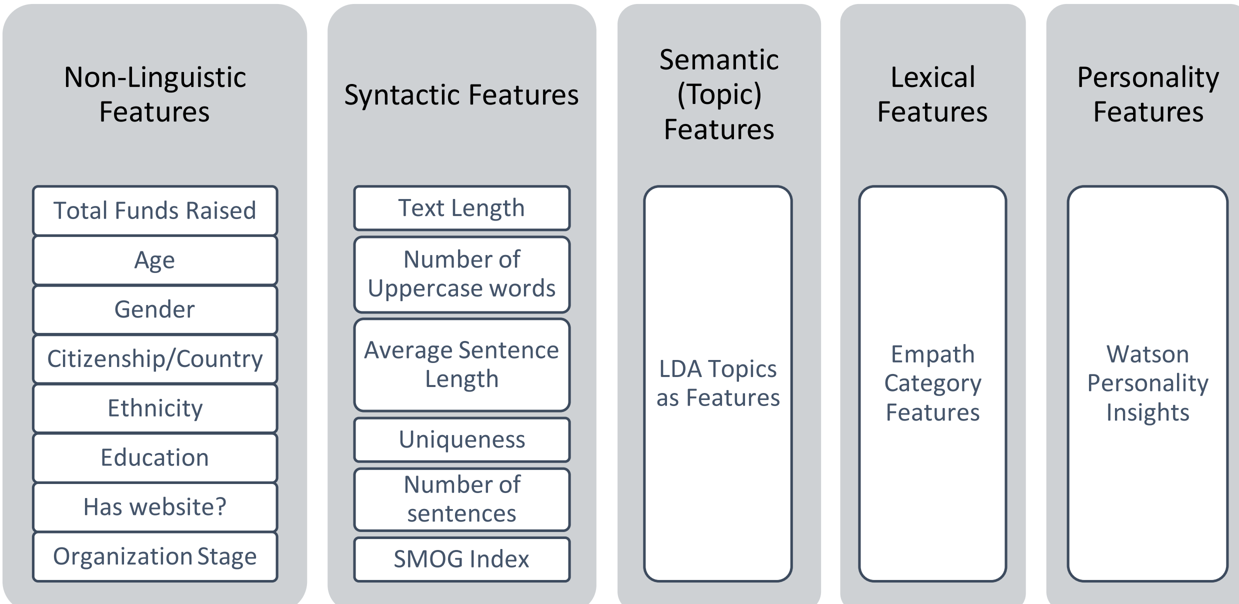


Figure 2. We divide the features we employ in this study in four classes covering non-linguistic, syntactic, semantic, lexical, and personality features.

Given that one of our goals is predicting successful applicants, here we describe several application characteristics—including information about the applicants—and the set of features capturing them; overviewed in Figure 2, and which we detail below.

2.1.1 Non-Linguistic Features

The first class of features covers Echoing Green applicants' responses to categorical questions about demographics, education, organization funding, online presence (e.g., having a website), and many others. In contrast to the other classes of features that we describe next, we used these features as provided with basic pre-processing.

2.2.2 Linguistic Features

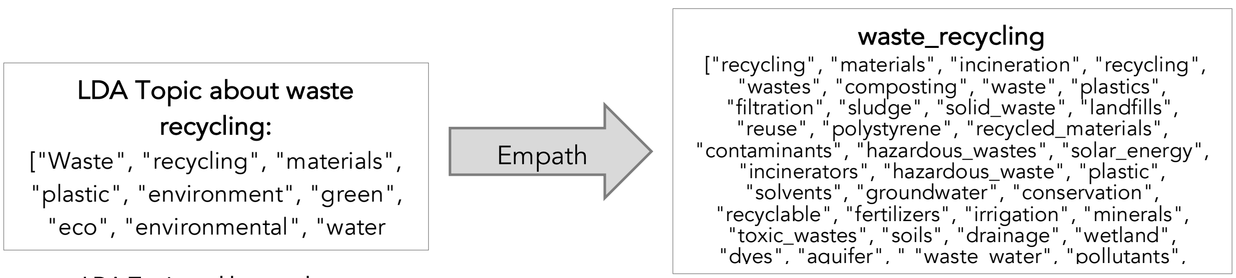
Syntactic Features: To characterize the way an application is written, we use a number of simple text processing approaches. For instance, to estimate the how difficult it is to understand the text written for an application we used a known measure of readability, the SMOG index.[2] To measure the distinctiveness of an application with respect to the rest of the entire application

[2] McLaughlin, G. Harry (May 1969). "SMOG Grading — a New Readability Formula" (PDF). *Journal of Reading*. 12 (8): 639–646. Retrieved 2016-12-07.

pool in each cycle, we use the sum of TF-IDF values of all terms which we refer to as the *uniqueness score*.[3]

Semantic Features: To explore if the topic of an application might be predictive of its' success, we use a popular topic modeling technique Latent Dirichlet Allocation (LDA)[4] to infer the topics on four categories of textual answers—describing the applicant, the problem area, the idea, and the organization structure. All text was pre-processed before applying LDA by removing stop-words and words with low frequency (<5).

Lexical Features: Further, we use the Empath tool to generate lexical categories and use the proportion of word in each category as features. Empath is a tool for analyzing text across lexical categories (similar to LIWC), and also generating new lexical categories to use for an analysis[5]. To account for the categories that the applicants talk about in the application and are not covered in the existing 194 Empath categories, we generate our own categories using the top 10 keywords from each of the topics obtained by applying LDA. E.g.-

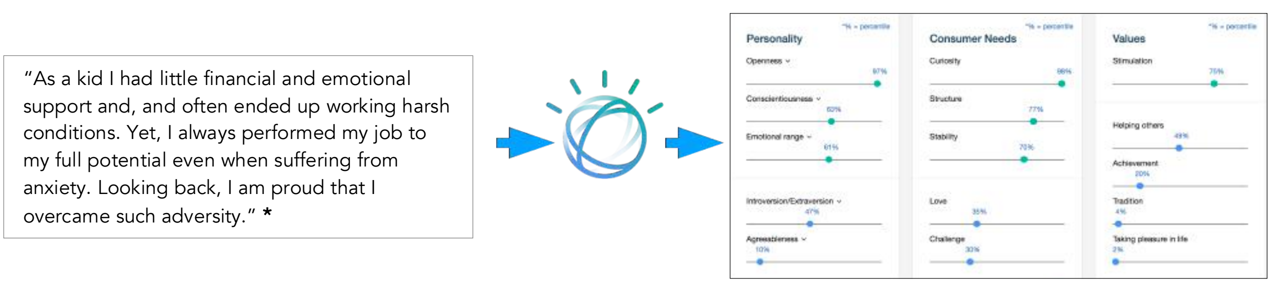


LDA Topic and keywords

Empath Category

Example of how top keywords from LDA generated topic are used to create custom Empath categories

Personality Features: Finally, we use IBM Watson's Personality Insights API to gain insights about the applicants' personality from text they write, which we also use as features. Watson's Personality Insights takes raw text as input, performs linguistic analytics and infers the user's personality by providing scores across the Big Five OCEAN traits, needs and values[6]. Example:



Example of how Watson API converts raw text into Personality Scores

---

[3] Hsu, Chiao-Fang, Elham Khabiri, and James Caverlee. "Ranking comments on the social web." Computational Science and Engineering, 2009. CSE'09. International Conference on. Vol. 4. IEEE, 2009.

[4] Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." *Journal of machine Learning research* 3.Jan (2003): 993-1022.

[5] Fast, Ethan, Binbin Chen, and Michael S. Bernstein. "Empath: Understanding topic signals in large-scale text." *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2016.

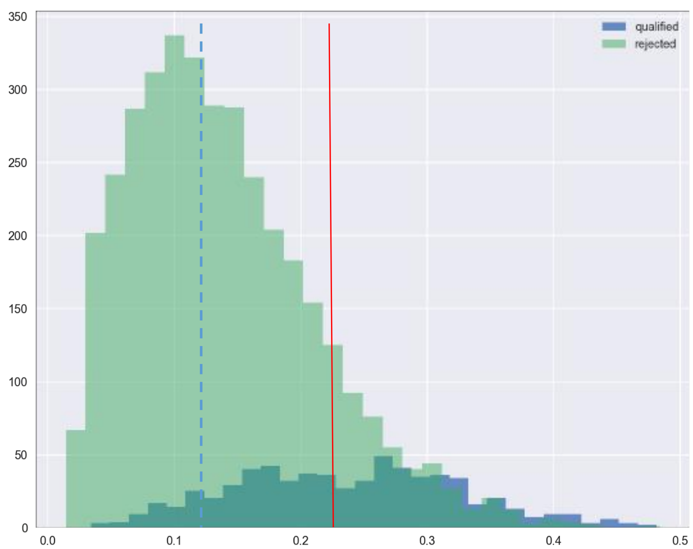
[6] https://www.ibm.com/watson/services/personality-insights/

*NOTE: This text does not belong to any application submitted to Echoing Green, and it is provided for explanatory purposes only.

2.3 AUTOMATED CLASSIFICATION OF APPLICATIONS

Using the set of features we described above, we train a random forest classifier to separate the applications into rejected or accepted for the phase 1 of the application process. The model assigns to each application a predicted probability score indicating its likelihood to qualify through this first phase. This allows us to, then, select the appropriate value for the probability threshold—the threshold over which all applications are predicted to be successful—to explore a range of precision and recall operating points that satisfy different classification objectives.

For instance, as the confusion matrix in Table 2A shows, with a probability threshold of 0.225, we correctly reject 78% of the applications with a false positive rate of about 8.2%, while 7% of applications are correctly accepted with a false negative rate of about 53% (equivalent to rejecting successful candidates in 8% of all observations). To minimize the false negative rate while maintaining a relatively high rejection rate, we can adjust the probability threshold. For instance, if we set the probability threshold to 0.125, we can successfully reject 53% of the applications while reducing the false negative rate to about 20%.



Confusion Matrix		
	Rejected ℗	Qualified ℗
Rejected	78%	7%
Qualified	8%	7%

Table 2a) Threshold = 0.225

Confusion Matrix		
	Rejected ℗	Qualified ℗
Rejected	53%	32%
Qualified	3%	12%

Table 2b) Threshold = 0.125

Figure 4. Probability Distribution across Applications; in the confusion matrices the percentages sum up to 100% .

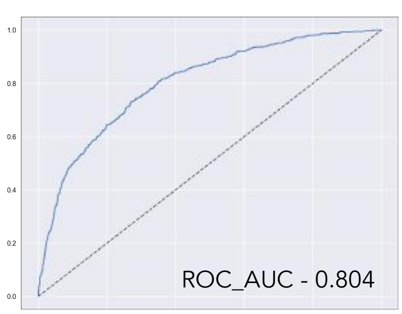


Fig 5. Receiver operating characteristic

Further, we also measure the Area Under the ROC Curve (AUC). We obtain a value of 0.80 which indicates that our classification model does a good job at separating the applications into the two distinct classes, and indicates that our feature set is predictive of an application success.

3. OBSERVATIONS AND RESULTS

Next we briefly overview part of our results, where we first describe the features our classification model identified as most predictive. We then discuss prevalent topics in the applications, and how they evolve over time.

3.1 What Features Are Most Predictive of Applications Qualifying in Phase 1?

The table below lists the top 25 most predictive features in our classification model:

Feature Type (Count)	Feature Name
Non-Textual (2)	Total Funds Raised, Re-applicant
Syntactic (8)	Number of Words – Applicant, Idea (2); Number of Uppercase Words - Applicant; Text Uniqueness – Applicant, Organization, Idea, Problem (4); Number of Sentences - Idea
Semantic (8)	Social Business and Team Experience, Online Platform / Data, Legal Organization/Rights, Water/Clean Energy, Women/Children, Local/Community Work, Business Venture/Project, Youth/Society
Lexical (1)	Helping People category
Personality (6)	Self-transcendence, Curiosity, Immoderation, Openness to Change, Self-Discipline, Neuroticism

3.2 What Are the Popular Themes and Topics That the Applications Cover?

To explore this question, we start by visualizing the word tokens from qualified applications (see Fig. 6), which we scaled by their TF-IDF values; finding that the most prominent words used in the applications represent the three broad categories of the Echoing Green Fellowship – Climate, Global, and Black Male Achievement.

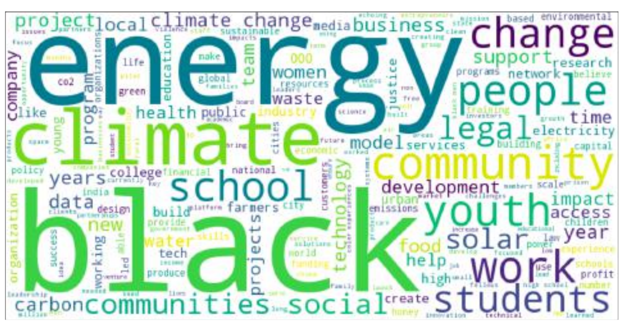


Fig 6. Word Cloud showing the distribution of words used by Phase 1 Qualified Applicants

Text Category	LDA Topics
Applicant Text	Development Project, Rural Food, Healthcare, Black Youth, Social Technology, Education, Vague Text
Problem Area Text	Climate Change, Clean Energy, Community Impact, Health, Education, Food & Agriculture, Poverty/Youth Training, Housing, Black Youth, Women/Children
Idea/Solution Text	Education, Food Production, Youth Community, Legal/ Organization Rights, Children, Waste Recycling, Women/Children Health, Health Care, Online Platform/Data, Clean Energy, Black Community, Sustainable Development, Social Business Venture
Organization Structure Text	Profit/Non-Profit, Experience, Mission, Role, Problem Area, Investors/Finance, Vision, Responsibilities

Table 4 – Topics extracted using Latent Dirichlet Allocation Model

We, then, used LDA to extract topics across applications from the detailed textual responses describing (1) the applicants, (2) the addressed problem area, (3) their proposed solution, and (4) the organization structure (as mentioned in Section 2.2). Specifically, the topics extracted from the *applicant description* text characterize their background, their experience, and their area of interest. Further, the topics extracted from the text describing the *problem area* highlight various domains such as clean energy, education, and poverty and youth training. Similarly, the topics extracted from the *proposed solutions* indicate the domain of the solution being proposed such as food production, waste recycling, and legal rights. Finally, the topics related to the organization structure cover discussions about the different roles that the applicant holds in the organization, their responsibilities, the revenue model of the organization, among others. The topics identified by LDA for each of these four categories are listed in Table 4.

Further, using the topics distilled by LDA, we analyzed the trends in the topic distribution across different stages of the application evaluation process, observing that certain topics are over-represented after the first qualification phase (i.e., semi-finalists) with respect to other topics. For instance, in Fig. 5, we see that proposed solutions related to *green technology and energy* have seen a growth in the number of applications, while remaining consistently over-represented across years among semi-finalists with respect to the initial pool of applications. In contrast, the number of applications concerning *youth employment* has decreased, being also increasingly under-represented, over the years, in later evaluation phases. Interestingly, though the proportion of applications addressing various issues related to the *Black community* have seen a decrease in the number of applications, they have been increasingly better represented among the semi-finalists. This may be an artefact of a specific focus on applications addressing this problem area, including through the introduction of a dedicated fellowship category on Black Males Achievement.

Finally, a few examples of the differences in topic prevalence between qualified and rejected applications can be seen in Fig. 6, where applications discussing prior team experience and social impact, addressing problems in the water and energy domain, or proposing online and data-based solutions appear more likely to become a semifinalist, i.e., qualify through Phase 1.

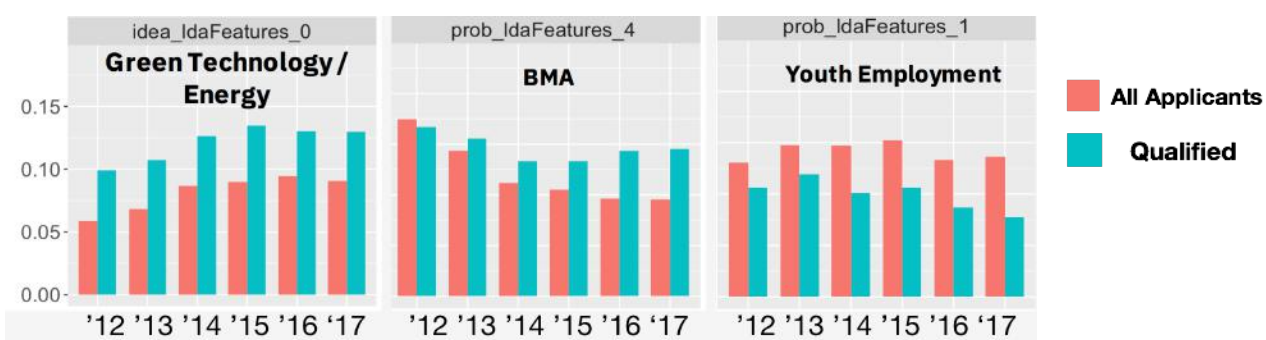
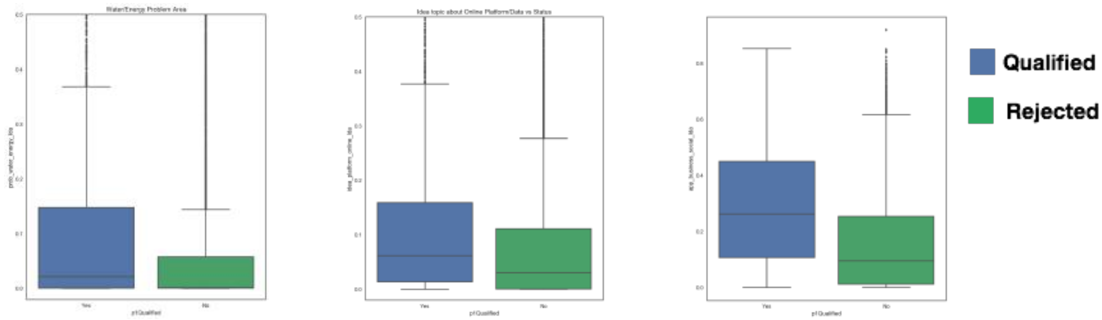


Fig 5. Trend of topic distribution

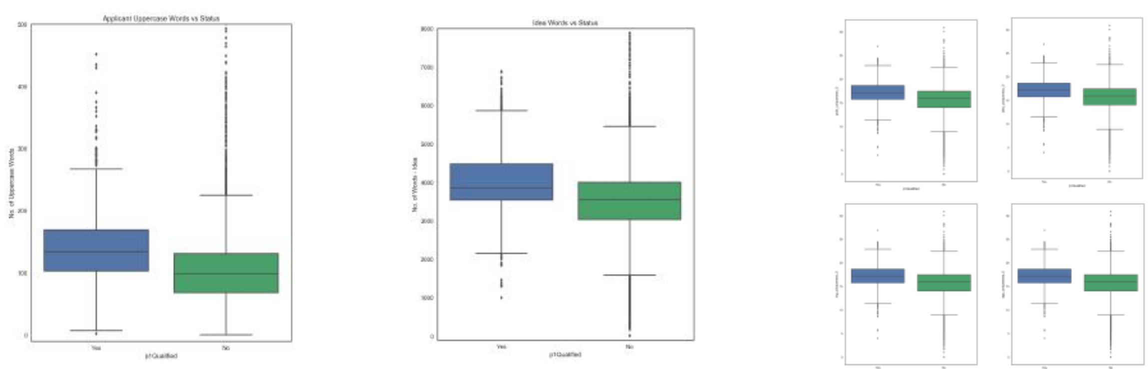


Water/Clean Energy Problem Area | Idea about online platform / Data | Applicant topic about social impact, team experience

Fig **6**. Box plots contrasting distribution of topic features across qualified and rejected applicants.

3.3 Are Successful Applications Written Differently?

By looking at the differences in the distribution of syntactic features between the applications rejected in the first qualification phase and the semi-finalists, we notice that various patterns in how an application is written are held at higher rates by successful applications. We see, for instance (see Fig. 7), that among the semi-finalists there are more applications that are described more extensively (i.e., more text), include a higher number of capitalized words (e.g., likely due to using more acronyms), and use more distinctive language (resulting in a higher uniqueness score, see Section 2.2/syntactic features).



Number of uppercase letters in applicant text | Number of Words in Idea text | Application text uniqueness

Qualified **Rejected**

Fig **7**. Box plots contrasting distribution of syntactic features across qualified and rejected

3.3 What Personality Traits Are More Prevalent Among Fellows?

As per their mission, Echoing Green aims to find the right applicants, rather than the right ideas. Therefore, we also explore key personality characteristics of successful applicants as inferred from the textual content in their applications using Watson Personality API (see Section 2.2/Personality features). In Fig. 8, we observe higher personality scores among successful applications on neuroticism (related to mood change), extraversion, and self-discipline, as well as lower scores on immoderation (lacking restraint).

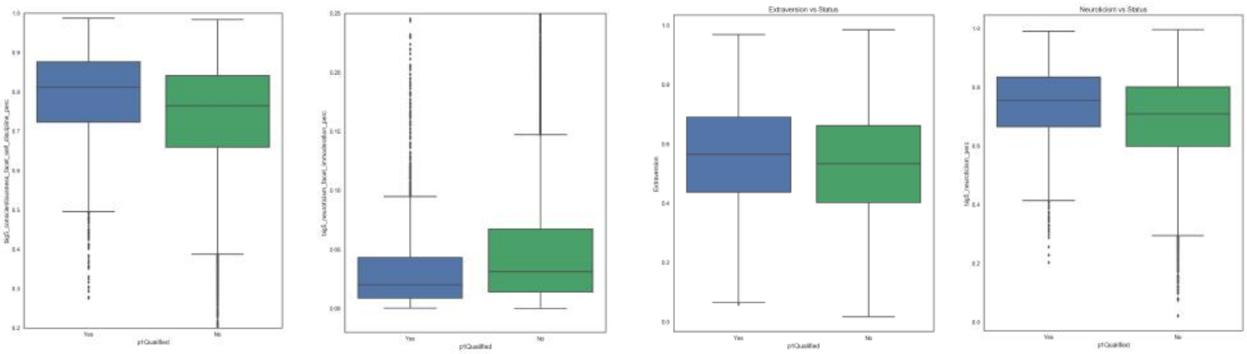


Fig **8**. Box plots contrasting the distribution of personality features across qualified and rejected

3.5 Are There Predictive Patterns in the Answers to Other, Non-Textual Questions?

Finally, we also considered the distribution of answers to non-textual questions (e.g., our list of non-textual features), to see if certain answers are more prevalent among the applications qualified through phase 1 of the evaluation process. We observe (Fig. 9) that more money was raised for successful applications, than for unsuccessful ones. The amount of money raised could be an indicator of an applicant's resource magnetism, which is something **Echoing Green** values. Another factor more prevalent among successful applications is whether the applicant has submitted a fellowship application in the past or not—in Fig. 9b) we see that resubmitting the same idea makes an application twice more likely to get accepted. Other factors such as having a website for the organization, are also more common among successful applications.

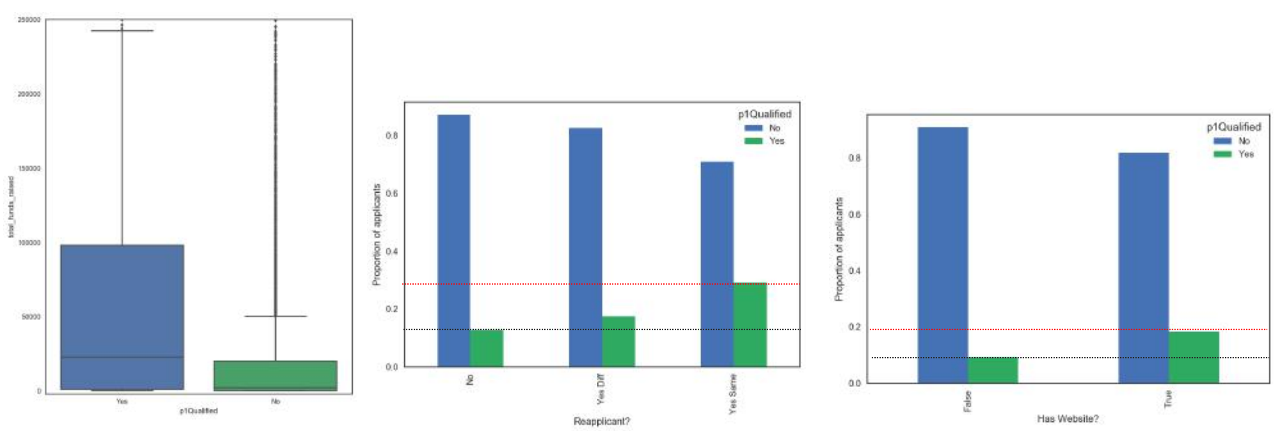


Fig 9a. Total Funds Raised

Fig 9b. Re-applicant

Fig 9c. Has Website?

Fig **9**. Distribution of non-linguistic features across qualified and rejected applicants.

4 CONCLUSIONS

In this study, we used natural language processing to engineer features from free text in application data from last six years. Our main overarching goals were to explore what type of characteristics are held more commonly among applications making it through initial evaluation round, and the extent to which we can train a classification model that is able to distill successful applications from a larger pool. We find that by probing a range of precision and recall operating points we can satisfy different classification objectives. Further, by exploring the differences in the distribution of a variety of features, we learned that successful applications tend to be longer (in terms of text length), more distinctive (or unique), and they also tend to, for instance, more often address problems from certain domain areas such as climate and energy. We also observed that certain personality traits seem to be more associated with applicants making it through the final evaluation phases. The ability to look at such common characteristics enables organization such as Echoing Green to reflect on whether they represent institutional perspectives on social change. The overall insights about the application pool can further be used to improve the reviewer training process and to inform reviewers' evaluation criteria, while the automated classification can be used to highlight specific applications for a more focused reviewing.