

variety of social enterprises globally—annually vets around 3,000 applications and selects just over 1 percent of them for a two-year fellowship that includes both financial, as well as leadership development support. Applications for their Fellowship contain comprehensive information about both the applicants and their bold ideas for social change.

The rich textual data contained in years worth of fellowship applications provides a unique opportunity to analyze the pool of applications and perhaps gain insights into trends and what makes applicants successful in achieving social impact. We believe such insights could help the broader community of social entrepreneurs to better direct their efforts and magnify the collective impact they can achieve. Some questions to help address this need are; What do the applications focus on and how did this focus changed over time? What factors differentiate successful applications? Do they contain cues about what it takes to achieve [social impact](#)? Do individuals of different demographics and traits tend to focus on different topics?

To explore such questions, this summer, a group from the IBM Science for Social Good Program teamed with Echoing Green to use machine learning and natural language processing techniques to extract explanatory cues from this unique collection of anonymized application data. Though much of Echoing Green's work is to help dismantle barriers to opportunity for its Fellows, they recognize the importance of regularly and rigorously evaluating their own search and selection processes as a way to help dismantle structural barriers to entry for emerging entrepreneurs across the globe.

The effort was led by our IBM Social Good Summer Fellow, Aditya Garg – a graduate student at Columbia University – and includes several data science researchers from IBM Research. Our team's focus was on distilling the traits that are predictive of successful applications and to

run an exploratory analysis to identify trends in the data. Some initial results of the project are below.



Figure 2: On top, we see the prevalence across years and among the entire pool of applications (red) versus among the finalist applications (blue) of the green technology & energy topic within answers related to the solutions, social impact & teamwork topic (answers about the applicant), climate topic (answers about the application domain). On the bottom, we see the most prevalent terms for each of these topics.

What do the applications focus on and how this focus changed over time? Having data spanning several years was key in distilling patterns about interests and their temporal variation, yet this also came with challenges, as the application questionnaire has evolved over the years, balancing between questions focusing on various aspects related to social entrepreneurs and their ideas. To work around these variations, we categorized the text answers to questions falling in four categories about the applicant, their organization, the application domain, and their

proposed solutions; allowing us to track trends across them for both the original pool of applications and for those applications that progressed in the evaluation process.

By doing so, and using an unsupervised topic modelling approach, we could observe patterns in under-representation and over-representation of topics among those applications that qualified to later phases in Echoing Green's evaluation process. In the figures above we see examples of topics that were both over-represented in later evaluation phases and whose prevalence among the initial application pool has increased over the years. Specifically, our results indicate that talking about "impact" and "team work" has both increased and makes applicants more likely to qualify through the evaluation phases, and that there is an increasing interest in applicants proposing both [climate change](#) as a problem area, and green technology and energy as part of the solution.

Patterns in how the application are written provide subtle cues that are predictive of how they fare in the fellowship evaluation process. We know from past research that there are links between the use of language and cognitive styles, organizational structure, and behavior. Tools like IBM Watson Personality Insights API extract such cues about Big Five OCEAN traits, needs, and values from textual content through linguistic analytics, helping us to partially operationalize the applicant criteria that Echoing Green values (such as purpose, resilience, or leadership.) For instance, [applications](#) obtaining a higher score for immoderation (indicating an orientation towards short-term pleasures and rewards, rather than those that require a longer-term commitment) tend to be rejected earlier in the evaluation process, while those with higher scores for self-discipline (indicating persistence in completing difficult, unpleasant tasks) tend to progress in the evaluation process.

These are only a handful of examples from this summer, and we know

further work is needed to translate the insights into actual actionable guidelines for evaluation purposes, or to support entrepreneurs.

However, we believe this approach holds great potential, as we think the ability to look at common characteristics enables organizations to reflect on whether they represent institutional perspectives on social change, and how the overall insights about the application pool can further be used to inform their evaluation criteria.

Echoing Green has already begun to discuss these possibilities, as Liza Mueller, Director of Operations & Knowledge Management reports, "Our internal conversations about the findings have focused around how we can better respond to and inform others on trends in the field, support prospective applicants in clearly understanding and responding to our application questions (regardless of location, education level, etcetera), as well as adapting evaluator training methodologies so that our process continues to both allow us to spot the best-in-class [social change](#) talent and serve as a model for others. We believe talent is equally distributed but opportunity is not, and this robust analysis from the IBM Science for Social Good Program will support us in further building a diverse, innovative, and impactful community."

Provided by IBM

Citation: Demystifying social entrepreneurship—a data-driven approach (2017, September 28) retrieved 26 April 2024 from

<https://phys.org/news/2017-09-demystifying-social-entrepreneurship-a-data-driven-approach.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
