SIGKDD Impact Program 2018

Rok Sosič
Department of Computer Science
Stanford University
rok@cs.stanford.edu

Rayid Ghani
Center for Data Science and Public Policy
University of Chicago
rayid@uchicago.edu

Jure Leskovec
Department of Computer Science
Stanford University
jure@cs.stanford.edu

1. PROGRAM DESCRIPTION

The SIGKDD Impact Program was established in 2017. As the SIGKDD community has expanded its reach dramatically over the past few years and the KDD conference has grown into a major global event, the aim for the social impact program is to focus the power of the community towards a broader positive societal impact.

The goal of this program was to support projects that promote data science, increase its impact on society, and help the data science community. The project duration was limited to one year. The total amount of funding provided by the SIGKDD Impact Program in 2018 was $250k, given as unrestricted gifts.

Projects funded in 2018 were required to present a mid-project update of their work and outcomes at the KDD 2018 conference in London in August 2018.

2. SELECTION CRITERIA

The SIGKDD Impact Program was looking to fund projects that have the potential for maximum impact on the society and expand outreach of data science thereby strengthening the community. Sample topics provided in th call for project proposals to guide proposals were:

- Enhance data science community engagement;
- Expand outreach of data science;
- Increase diversity and participation in data science;
- Increase societal impact of data science;
- Influence public policy through data science.

The program was not intended for research projects, but instead focused on projects that benefit the community and the society. The call for projects was open to all countries, and participants were expected to be affiliated with an institution with a reputable standing in the KDD community. Funds were intended for non-profit activities where no alternative funding is readily available.

3. SELECTION PROCESS

The call for proposals resulted in a broad participation with 75 submissions from 20 countries, covering a wide range of topics. The following KDD members served on the selection committee: Rayid Ghani, Jingrui He, Jure Leskovec, Jian Pei, Claudia Perlich, and Mohammed Zaki. Seven proposals from five countries were awarded funding according to the program selection criteria.

The selection committee members selected the top proposals as follows: Each submission was evaluated by three randomly assigned judges, with each committee member judging half of the submissions. Each judge produced a ranked list of top 10 submissions from their assigned list of proposals. These six individual lists were merged into two unified rankings, applying two different methods. One method used judgments from all three judges, the second method only the two best judgments. The two resulting unified rankings were then discussed by the committee members, which selected the best seven proposals for funding.

4. SELECTED PROPOSALS

Below are brief descriptions of the projects that were selected for the 2018 SIGKDD Impact Program.

Training the Next Generation of African Data Scientists, Kenya,
Project leaders: Isaac Markus and Audrey Cheng,
Cover organizations: IBM Research Africa and the Moringa School.
The project sought to establish a part-time data science course available to Kenyan students from varied educational, technical and economic backgrounds, with the aim of expanding their employability as data analysts and data scientists.

Budding Data Scientists Hackathon, Singapore,
Project leader: Hui Xiang Chua,
Cover organization: Hwa Chong Institution.
The project aimed to motivate upper secondary school students to develop an interest in data science and to help them tackling social challenges of their interest using data science, with a possibility of improving the data maturity within Voluntary Welfare Organisations, i.e. non-profit organisation that provides welfare services or services that benefit the community at large.

Women Data Science Leaders in Russia, Russia,
Project leaders: Alena Suvorova and Valeria A. Ivaniushina,
Cover organization: National Research University Higher School of Economics.
The project aimed to increase gender diversity and women's participation in the Russian data science community by developing online courses and video materials that present female role models for female students, in order to change the stereotypes that affect the perception of the data science field.

Community Perspectives on the Use of Algorithms by Government, New Zealand,
Project leader: Rhema Vaithianathan,
Cover organization: Centre for Social Data Analytics, Auckland.
This pilot study sought to understand how comfortable different communities are about the use of algorithms and machine learning tools in government. The study focused on people who are most likely to be subject to or affected by algorithms, for example those who are on public benefits, have child welfare referrals, or have been through the criminal justice system.

Empowering Community Members to Understand and Use Big Data and Big Data Methods, United States,
Project leader: Lisa Schilling,
Cover organization: University of Colorado, School of Medicine, Division of General Internal Medicine.
The objective of this proposal was to create reusable education materials for engaged citizens and community leaders regarding big data and big data methods by understanding and developing content around the specific use case of a community initiative to improve the mental health.

Data Science for Social Good Marketplace, United States,
Project leaders: Rayid Ghani and Mohak Shah,
Cover organization: Center for Data Science and Public Policy, University of Chicago.
The project proposed to build an online marketplace where governments and non-profits can post their data science project needs, and where volunteers with expertise in data science project scoping can ask questions and turn those needs into well-defined problems, and where data scientists can team up to solve those problems collaboratively. The community can serve as a marketplace, a discussion board, and a collaborative workspace. The marketplace is now active[1].

Summer Academy in Data Science for High School Students, United States,
Project leaders: Bamshad Mobasher, Lucia Dettori, Raffaella Settimi, and Daniela Raicu,
Cover organisation: DePaul University.
The project proposed to develop a week-long summer program to introduce high school students to the field of data science through a series of hands-on activities. The aim was to help students build confidence in their problem-solving skills through data science by introducing them to the full data science cycle from data cleaning and visualization to model building and validation.

5. RESULTS

The projects were asked to provide a presentation at the SIGKDD KDD 2018 Conference in London in August 2018. These presentations were given at the KDD Workshop on Social Impact at the conference[2].

We are excited to be able to share written reports from four out of the seven projects in this issue of SIGKDD Explorations. The projects providing their reports are: Data Science Summer Academy for Chicago Public School Students, Women Data Science Leaders in Russia, Empowering Community Members to Understand and Use Big Data and Big Data Methods, and Budding Data Scientists Hackathon. Additional reports are scheduled to appear in the next issue of SIGKDD Explorations.

6. ACKNOWLEDGMENTS

The authors would like to thank the selection committee members for their service, and the editor of SIGKDD Explorations, Hanghang Tong, for his initiative to publish the reports and make them available to the wider KDD community.

[1] https://www.solveforgood.org

[2] https://dssg.uchicago.edu/kddsocialimpact/