Big Data CRUX Program: Development of a Collaborative Resource and Understanding Exchange

Call for Applications

The Northeast Big Data Innovation Hub (https://nebigdatahub.org) is pleased to invite applications for the Big Data CRUX Program. Six motivated PhD students from institutions across the Northeastern United States will be selected to join a multi-disciplinary project, culminating in the development and implementation of an open source platform that will spark collaboration and solution-building by connecting real-world challenges with shared data resources and expertise. The program is led by Andreas Müller, Lecturer at the Data Science Institute at Columbia University, and René Bastón, Executive Director of the Northeast Big Data Innovation Hub.

About the Northeast Big Data Innovation Hub

The Northeast Big Data Innovation Hub is one of four regional Hubs forming a national network of academic, industry, and government partners supported by the National Science Foundation (NSF), with a mission to stimulate innovations enabled by Big Data and data science across the nation. Based at Columbia University in the City of New York and representing data-driven projects and institutions throughout our region, the Northeast Hub identifies high-priority societal challenges and builds public-private, multi-sector partnerships and consortia to address them with data-driven solutions.

About CRUX

The CRUX data sharing platform is a project developed at the NEBDIH, representing a community-driven approach to accelerating data-driven innovation. CRUX will provide an open online platform for sharing data sets together with well-defined, domain-specific tasks derived from target stakeholders. The goal is to provide real-world applications and data sets to data scientists, while providing subject matter experts from academia, corporate and government sectors access to Big Data expertise.

Program Activities and Timeline

Up to six doctoral candidates at institutions across the Northeastern United States will receive funding as part of this effort. Participants will meet with program leadership at the Northeast Big Data Hub's headquarters at Columbia University in the City of New York, for a series of platform development meetings and career development activities (further described below). After this planning stage, program participants will return home to work as ambassadors engaging leaders in their fields of study, assessing their data science needs and collecting use cases and

data sets for the platform. At the end of this research period, students will reconvene to share their results. Students will be mentored by the program leaders René Bastón and Andreas Müller, and will receive a stipend and travel funding for their participation.

Career Development Module

Beyond platform development, the program will afford its participants immersive career development opportunities**.** The first is through participation in a workshop on communication, data narrative and storytelling. Participants will also be connected key stakeholders from industry, academia and government entities in their respective areas of study, to assess data science needs and collect use cases in the Northeastern United States. These opportunities will provide participants with experience in requirement gathering, project planning and communications.

Desired Skills

Candidates must be matriculated Ph.D. students at a university in the Northeastern United States (Connecticut, New Hampshire, Maine, Massachusetts, New Jersey, New York, Pennsylvania, Rhode Island, Vermont). Experience with data-driven research and data science methods are advantageous. The following areas of expertise will strengthen a candidate's application:

Nontechnical Expertise:

- Candidates should have a background/substantial understanding of one of the following areas: cybersecurity risk/systemic risk; transportation; or health. An understanding of challenges in these areas revolving around the use of data or requiring the use of data to solve is preferred. An understanding of data science approaches is desirable.

Technical Expertise:

The following skill sets may benefit a candidate's application but are not required to apply:

- Python
- Python for web development
- Machine Learning

Application Process

Please complete the application form (https://bit.ly/ApplyCrux) to send your application including cover letter, CV and optional letter of reference. Submission deadline is Monday, November 20th at 5 pm Eastern Time. Decisions will be issued by mid-December.

Contact Katie Naum at ken2115@columbia.edu with any questions about the program.