COLUMBIA UNIVERSITY
Data Science Institute

Scikit Learn - Post Doctoral Opening

Postdoctoral opening at Columbia University Data Science Institute

The Columbia University Data Science Institute is seeking a postdoctoral researcher to work on a two-year project to improve software infrastructure for automatic machine learning based on the scikit-learn machine learning library.

This effort is part of the NSF Software Infrastructure for Sustained Innovation (SI2) program.

The postdoc will primarily work with PI **Andreas Müller**, but is encouraged to work with other researchers and students at the Data Science Institute. The initial term of this appointment will be for one year, and the start date will be on a mutually agreed upon date during in Fall 2017. The appointment is renewable for up to a total period of two years, subject to the usual standards for satisfactory performance, availability of funding and continued visa clearance (if applicable).

The NSF SI2 program is software-oriented, so the main output of this project will be software, largely as components of scikit-learn. While the goal is to publish novel methods and improvements on current system as research papers, this is secondary to the creation and publication of the software itself.

The project focuses on the development and benchmarking of good parameter search spaces for scikit-learn models, better preprocessing methods for scikit-learn, easier integration of scikit-learn with automatic machine learning packages, and creation of a flexible and easy-to-use meta-learning framework compatible with scikit-learn. The project is likely to include interactions with or contributions to the scikit-optimize, auto-sklearn and OpenML projects.

Qualifications:
- A PhD in Machine learning or a related field. In exceptional circumstances this requirement might be waived if you have extensive experience in machine learning and software development.
- A strong research record in machine learning or a closely related field.
- A strong mathematical background, in particular in linear algebra, optimization and probability theory.
- A broad understanding of different machine learning technologies.
- A track-record of high-quality software development, ideally as part of an open source project, Python and Cython preferred.

Please submit:
- A cover letter describing your relevant experience in machine learning and software development
- Your CV
- A link to your github account, or other major software written by yourself (ideally in collaboration)

Please send all application materials to DSI-SI2@columbia.edu.