

Scikit-learn developer

Scikit learn

We are looking for a junior developer in charge of supporting the development of the scikit-learn, a machine learning toolkit written in Python/C, and available at the following address : <http://scikit-learn.sourceforge.net/>. Scikit-learn provides generic tools to learn information for the classification of various kinds of data, such as images or texts. It is tightly associated to the scientific Python software suite (numpy/scipy) for which it aims at providing a complementary toolkit for machine learning (classification, clustering, dimension reduction, regression). There is an important focus on code quality (API consistency, code readability, tests, documentation and examples), and on efficiency, as the scikit-learn compares favorably to state-of-the-art modules developed in R in terms of computation time or memory requirements. Scikit-learn is currently developed by about 30 contributors, but the core developer team has been with the Parietal INRIA team at Saclay-Île-de-France since January 2010. The scikit-learn has recently become the reference machine learning library in Python. For the next 2 years, we propose to extend and enhance it without losing the strong API consistency of the tool.

Job description

The main goal of the junior developer will be to assure a high quality in the different contributions that are submitted for inclusion into the scikit, in close collaboration with the main developers. She/he will also evaluate the performance of the available code with respect to reference implementations, foster the communication between contributors by providing information and comments on the mailing list, while being involved in the organization of coding sprints gathering the developers (e.g. every 6 months). She/he will be in charge of diffusion and packaging work. In 2011-2013, the scikit should, among other things, include new functionalities for on-line learning, efficient model selection/evaluation and unsupervised data analysis; functionalities that the hired developer will assist in implementing in the scikit.

About the team

Parietal is a small team of computer scientists developing algorithms for data processing and modeling of brain images. It is embedded in Neurospin, a research institute dedicated to various aspects of brain imaging and grouping cognitive neuroscientists, medical doctors, imaging physicists and computer scientists. As a member of Parietal, the developer will benefit from the rich scientific environment of Neurospin. Parietal is part of INRIA, the French national computer science institute. As an engineer employed by INRIA, the developer will receive support and advice from the senior developers of the institute. Parietal is committed to using and developing scientific Python, and hosts core contributors to several scientific computing libraries in Python.

Required education profile

As this is a junior developer position, the candidate must have graduated from a master or equivalent diploma no more than one year before the start of the contract. We stress that we cannot hire candidates with more experience.

Required skills

- Programming in Python/C/C++
- Software quality (test-oriented development, version control [git], technical documentation)
- Unix/Linux environment
- Software design: impose a consistent API while integrating code from various contributors
- Knowledge of the open-source community
- Technical English
- If possible, some experience in statistical learning.

Salary

Gross salary 2500 euros per month

Duration

24 months, starting ca october 2011.

Contact

PARIETAL team, INRIA Saclay-Île de France
CEA Saclay, Bât 145, Point Courrier 156
F-91191 GIF/YVETTE, FRANCE
Bertrand Thirion, bertrand.thirion@inria.fr
Gaël Varoquaux, gael.varoquaux@normalesup.org