Launch of the **scikit-learn initiative**, a reference software library for machine learning

Inria is pleased to announce the launch of the scikit-learn initiative, a partnership with companies using scikit-learn. Its objectives are to support the development of this reference software: sustaining its high quality and adding new functionalities. Scikit-learn is a library in Python, an high-level programming language. It is dedicated to statistical learning (machine learning) and can be used as middleware, especially for prediction tasks.

**Ten years of research and development**

Initially launched in 2007 by members of the Python scientific community, the scikit-learn project had a new start in 2009 with the investment of Inria's Parietal team. To conduct research on brain imaging, the team needed a predictive modelling tool that integrated with the Python ecosystem. It then organised an open participatory development with the objective of building an open-source tool for statistical data analysis. Two years later, a first version was released.

Scikit-learn is now supported by a very large team of developers based in Paris, New York, Sydney and around the world. It is in the top three most popular machine learning software programs on GitHub.

**Ambitious objectives**

Clear objectives were set at the start of the scikit-learn project: covering reference machine-learning models with a high quality standard. So that the library could be easily used, the development team made sure that it was well packaged and wrote extensive documentation with concrete examples on the use of the tool. It also insisted that all methods be covered by a series of automatic tests that help ensure the quality of the code base over the long term.

The team now wants to push the library to new horizons while keeping the same ease of use and reliability.

**Retrieve complex data to make decisions**

Scikit-learn can process complex data (databases, texts and images) and classify them using state-of-the-art techniques for automated decision making.

Scikit-learn is open source and available under BSD license. A community of developers (inside and outside Inria) quickly formed, which made it possible to accelerate the development of the tool and foster many applications. A rich website (scikit-learn.org) provides a detailed introduction to the project and its applications. Scikit-learn is used by a large number of Web companies to predict user buying behaviour, offer product recommendations and detect trends and abusive behaviour (fraud, spam, etc.).

**Diversified fields of application**

One of scikit-learn's strong points is its generic nature, which ensures great versatility and diverse applications, such as:
- fighting against fraud and spam
- analyzing medical images
- prediction of user behaviour
- optimization of industrial and logistic processes.

For example, a general-public application as booking tourist venues uses machine-learning tools such as scikit-learn to automate tasks. With and understanding of the applications and the data they generate, a data scientist uses the library to build a powerful decision making system.
Scikit-learn is a constantly evolving, easy-to-use, effective and accessible statistical learning library for non-experts in data science. In the data mining stage, the user enters a few lines in an interactive interface and can immediately view the results of his analysis.

**The scikit-learn consortium**

To support and stimulate the scikit-learn ecosystem, a consortium of sponsors has been created with the support of the Inria Foundation. It will found engineers to ensure the quality of the project and the integration of new contributions, as well as the addition of ambitious new features. These efforts will be lead in close connection with scikit-learn’s vast community of users and developers. Both the foundation’s partners and the open-source community will be involved in defining the development priorities.

**Members of the scikit-learn consortium**

---

**About Inria:**

Inria, the French National Institute for computer science and applied mathematics, promotes “scientific excellence for technology transfer and society”. Graduates from the world’s top universities, Inria’s 2,400 employees rise to the challenges of digital sciences. With its open, agile model, Inria is able to explore original approaches with its partners in industry and academia and provide an efficient response to the multidisciplinary and application challenges of the digital transformation. Inria is the source of many innovations that add value and create jobs.

---

**Inria media contacts:**

**Headquarters**
Laurence Goussu - +33 (0)6 81 44 17 33
laurence.goussu@inria.fr

**Saclay – Ile-de-France Research center**
Magalie Quet - +33 (1) 77 57 80 27
magalie.quet@inria.fr