Scikit-learn: Machine Learning in Python

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, Édouard Duchesnay

Journal of Machine Learning Research 12 (2011) 2825-2830

Abstract

Scikit-learn is a Python module integrating a wide range of state-of-the-art machine learning algorithms for medium-scale supervised and unsupervised problems. This package focuses on bringing machine learning to non-specialists using a general-purpose high-level language. Emphasis is put on ease of use, performance, documentation, and API consistency.

1. Introduction

The scikit-learn project started in 2007 as a Google Summer of Code project by David Cournapeau. It has minimal dependencies and is distributed under the **simplified BSD license**, encouraging its use in both academic and commercial settings.

2. Project Vision

Scikit-learn aims to provide simple and efficient solutions to learning problems that are accessible to everybody. We favor a batteries-included approach: we rely on and require as few user-installed dependencies as possible.

3. Underlying Technologies

Scikit-learn is written primarily in Python, with some core algorithms implemented in Cython to achieve performance. It builds upon NumPy and SciPy.

4. Code Quality

The package is supported by a comprehensive user guide of **approximately 300 pages** including narrative documentation, class references, a tutorial, installation instructions, as well as **more than 60 examples**, some featuring real-world applications.

5. Authors

The authors of this paper number **16 contributors** total, representing institutions across France, Germany, and the United States.

6. Conclusion

Scikit-learn provides a simple and efficient solution for machine learning in Python, enabling both academic research and commercial applications through its permissive BSD license.