scikit-learn Contributor Experience Team

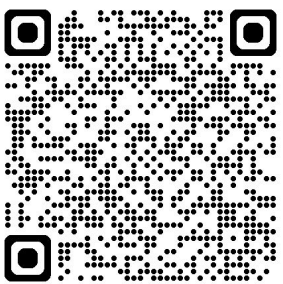
The role of a Contributor Experience Team in making contributions smoother

Supporting community member contributions via mentoring, scaffolding, and feedback

Reshama Shaikh @reshamas

Stefanie Senger @StefanieSenger

Virgil Chan @virchan



History of scikit-learn

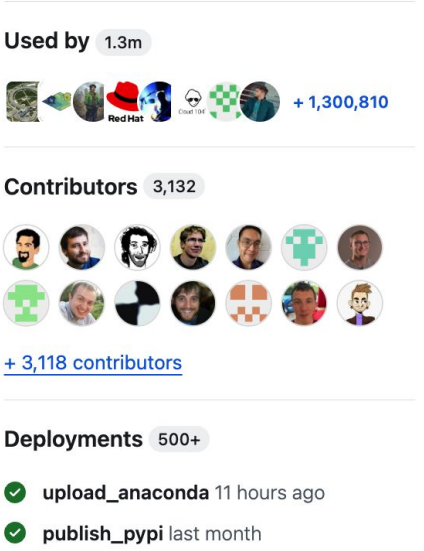
- Open source library for machine learning in Python.

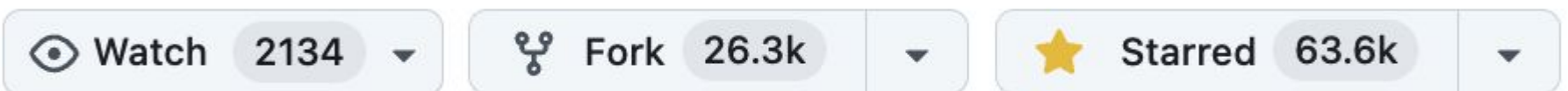
- Started in 2007 as a Google Summer of Code project by David Cournapeau
- Implement textbook algorithms

- In 2010, Fabian Pedregosa, Gael Varoquaux, Alexandre Gramfort and Vincent Michel of INRIA took leadership of the project
- Help people to do valid machine learning and data science as easy as possible.
- Developed and maintained by a community of volunteers, with major institutional support from INRIA and contributions from :probabl. engineers.

scikit-learn is a “large” and mature open source project

- Lots of code
- Lots of users
- Lots of contributors
- Lots of challenges





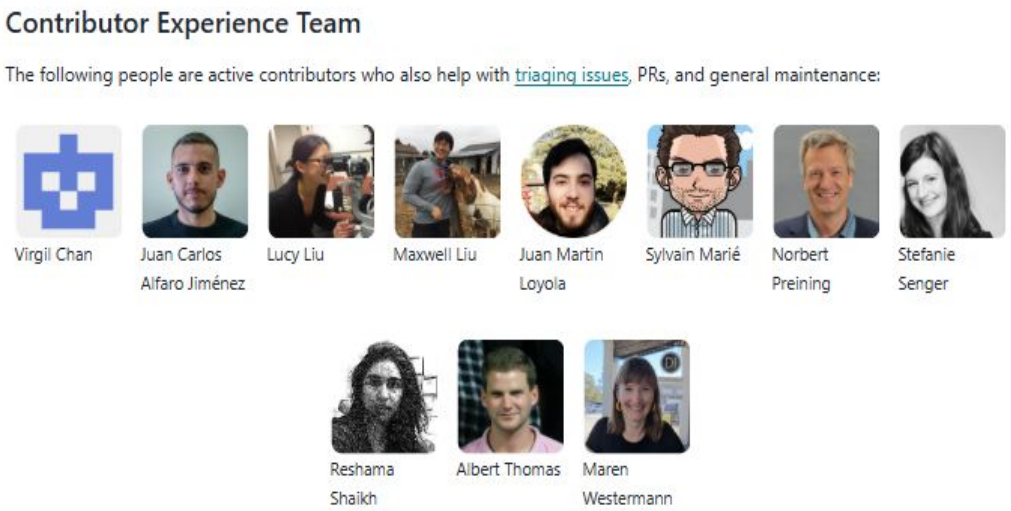
Governance

- Contributors
- Anyone who contributed to the project in any way — not just with code

- Core Contributors
- Maintainers
- Documentation Team
- Contributor Experience Team

Contributor Experience Team

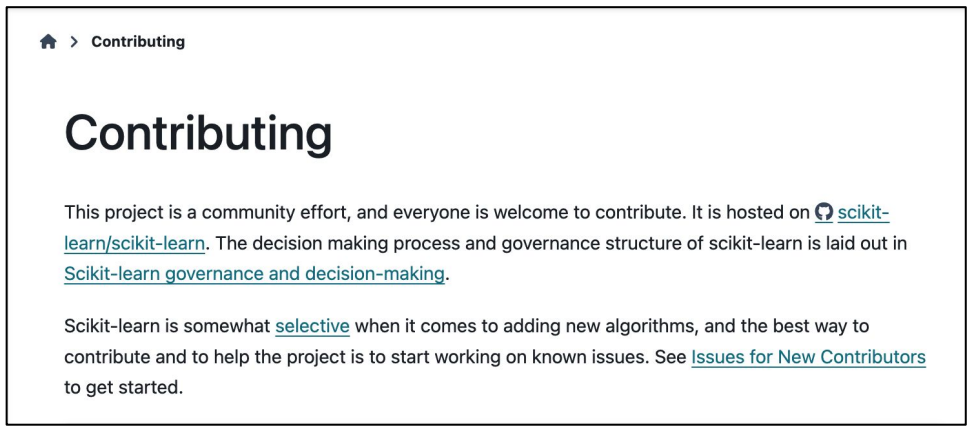
- Improve the overall experience for contributors
- Assist with triaging issues and pull requests
- Identify recurring challenges faced by contributors
- Propose and support improvements to address those challenges



Scaffolding

Resources for Contributors

CONTRIBUTING GUIDE



Contributing

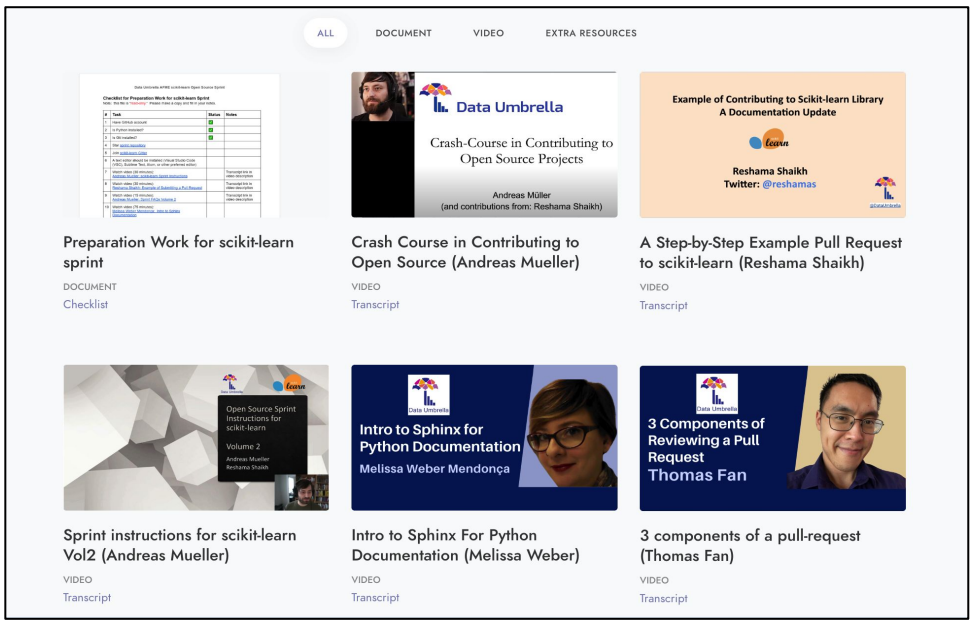
Contributing

This project is a community effort, and everyone is welcome to contribute. It is hosted on scikit-learn/scikit-learn. The decision making process and governance structure of scikit-learn is laid out in Scikit-learn governance and decision-making.

Scikit-learn is somewhat selective when it comes to adding new algorithms, and the best way to contribute and to help the project is to start working on known issues. See Issues for New Contributors to get started.

https://scikit-learn.org/stable/developers/contributing.html

VIDEOS



https://www.dataumbrella.org/open-source/contributing-to-scikit-learn

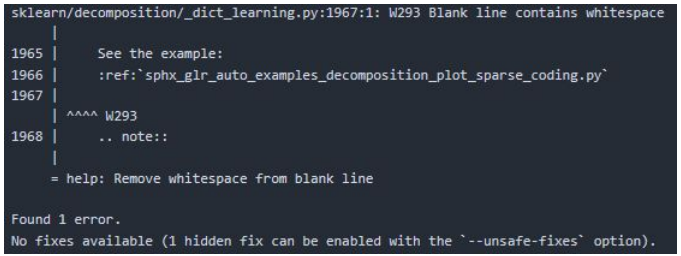
Good First Issue

scikit-learn/scikit-learn/30621

- Added example URLs to relevant sections in the API doc-strings and the User Guide
- Started from January 10, 2025
- Concluded in July 3, 2025
- Received over 80 pull requests

Common Beginner's Challenges

1. Linting Issues
    - Line length too long (over 79 characters, per PEP8)
    - Extra or missing blank lines
    - Incorrect indentation or spacing



<u>How the Team Helps:</u>

- Help contributors navigate CI/CD warning messages
- Encourage them to install pre-commit

Common Beginner's Challenges

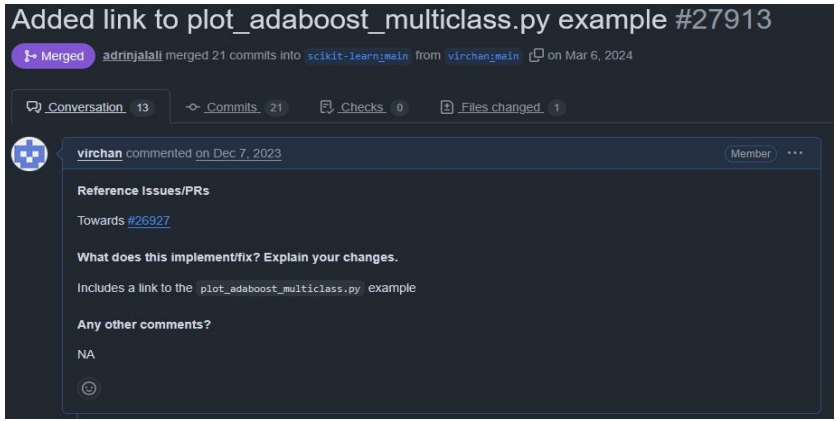
2. Navigating the Code Base
    - scikit-learn is a large project — 26100 Python files, 1174 reStructuredText, and more!
    - API guides and Examples are written in Python
    - User guide is written in reStructuredText
    - Difficult to identify the correct place to add example links for first-time contributors

<u>How the Team Helps:</u>

- Guide new contributors in locating relevant files and understanding the project's structure
- Leave constructive feedback to ensure examples are linked consistently and clearly

Example

scikit-learn/scikit-learn/27913



Hi @virchan! You added a link to the example in the example itself. The links need to be placed in the doc strings of the functions/ estimators and if applicable the user guide. For an example see PR #26877.

It's better to put the reference under the "Examples" section and write something like

```
For an example of how AdaBoost is used to fit a sequence of Decision Trees, refer to
:ref:`sphx_glr_auto_examples_ensemble_plot_adaboost_multiclass.py`.
```

It would also be good to leave a reference in the doc string of the DecisionTreeClassifier estimator under the "Notes" or "Examples" section.

Thank you for the update @virchan. Can you fix the linting issue? When you click on the details under "ci/circleci: lint" in the code checks you'll see the error messages.

Thank You!