

Probability 1.0

The community event for
scikit-learn® and
its open-source ecosystem



strategic
partners



:probabl.

Inria

HAZARDOUS



= Fairlearn

CHANEL

BNP PARIBAS
The bank for a changing world



Ecosystems, Not Just Code

Building Sustainable OSS Communities

25 Nov 2025



Stefanie Senger

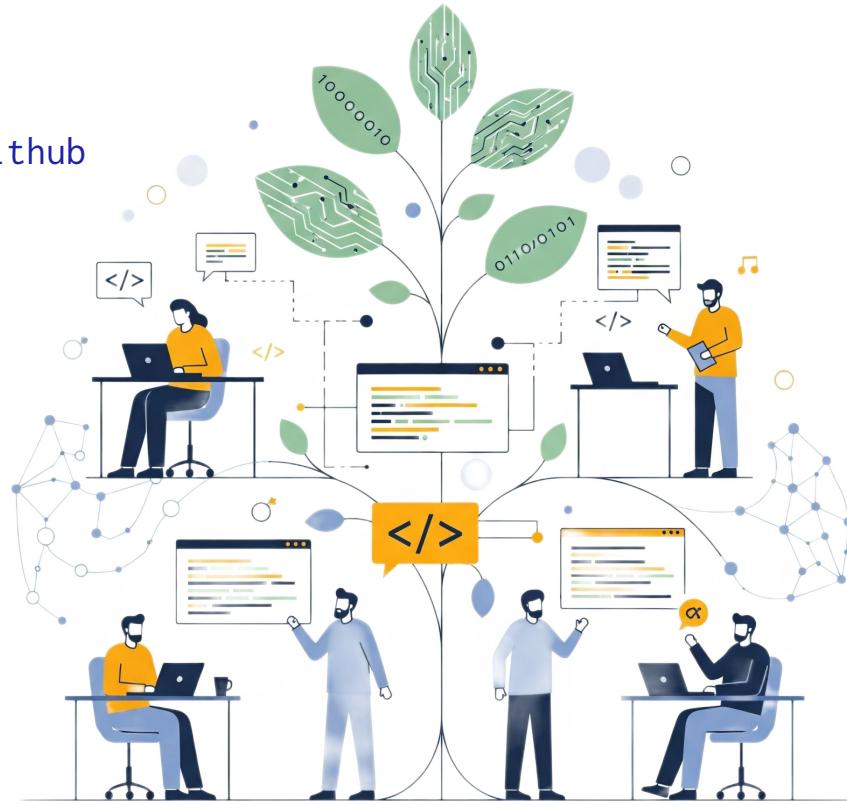
OSS Developer working at :probabl.

Probability 1.0

101 of Community Building

There is more to it than putting a project on github

- exchange and discussion
- coordination and decision making
- defining roles
- mentoring
- communication between strangers



101 of Community Building

exchange and discussion

 Clear current search query, filters, and sorts

- mature project (started 2007)
- 26100 Python files, 10000 pages of build documentation, and more
- total number of PRs: 20k
 - Open: +- 530
- total number of issues: 12k
 - Open: +- 1.6k
- a lot of topics already discussed

 525 Open  19,196 Closed

Author	Label	Assignee	Sort
 [MRG] GSoC 2014: Standard Extreme Learning Machines			 852
	Waiting for Reviewer		
	#3306 by IssamLaradji was closed on Feb 25, 2019		
 [MRG+2] Basic version of MICE Imputation			 703
	#8478 by sergeyf was merged on Apr 16, 2018		
 MRG: AdaBoost for regression and multi-class classification			 622
	New Feature		
	#522 by ndawe was merged on Feb 4, 2013		
 [MRG+1] Make cross-validation data independent + Reorganize grid_search, cross_validation and learning_curve into model_selection			 588
	#4294 by raghavrv was merged on Oct 23, 2015		

101 of Community Building

exchange and discussion

ENH Private Cython Submodule for Reduction over Pairwise Distances #20254

Edit  Code ▾

 Closed

jjerphan wants to merge 261 commits into [scikit-learn:main](#) from [jjerphan:pairwise_aggregation_cython](#) 

Conversation 428 Commits 261 Checks 0 Files changed +2,661 -446 

 jjerphan commented on Jun 11, 2021 Member ...
• edited

⚠ This PR has been superseded by #21462

What does this implement/fix? Explain your changes.

Aggregations over pairwise distances (argmin, argkmin, threshold filtering, cumulative sum, etc.) are one of the foundations used by many algorithms within scikit-learn.

Reviewers

 thomasjpfan	
 glemaitre	
 jeremiedbb	
 lorentzenchr	
 ogrisel	

Assignees

No one assign yourself	
--	---

101 of Community Building

exchange and discussion

ENH Private Cy
Reduction over
#20254

ENH Pairwise Distances ArgKmin #21462

Edit

< Code ▾

!! Closed

jjerphan wants to merge 322 commits into `scikit-learn:main` from `jjerphan:pairwise-distances-argkmin`

!! Closed

jjerphan wants to merge 261 com

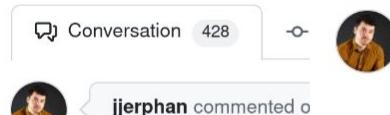
Conversation 107

Commits 250

Checks 0

Files changed 20

+1,939 -81



⚠ This PR has
[#21462](#)

What does this in
your changes.

Aggregations over pain
threshold filtering, cumu
foundations used by m:

jjerphan commented on Oct 26, 2021

Member

...

• edited

Reference Issues/PRs

Supersedes [#20254](#).

Context

Reductions over pairwise distances (argmin, argkmin, threshold filtering, cumulative sum, etc.) are one of the foundations used by many algorithms within scikit-learn.

Various strategies exist to compute them differently, including some using chunks (see [sklearn/metrics/pairwise.py](#)).

Reviewers

ogrisel

thomasjpfan

lorentzencchr

jeremiedbb

Assignees

No one — assign yourself

Labels

cython Enhancement High Priority
module:metrics module:neighbors

101 of Community Building

exchange and discussion

ENH Private Cy
Reduction over
#20254

ENH Pairwise

closed

jjerphan wants to merge 322 cor

closed

jjerphan wants to merge 261 com

Conversation 428



jjerphan commented on
edited

⚠ This PR has
[#21462](#)

What does this in
your changes.

Aggregations over pain
threshold filtering, cumu
foundations used by m

ENH Introduce

PairwiseDistancesReduction and
PairwiseDistancesArgKmin (feature
branch) #22134

Edit Code

Conversation 107



jjerphan commented

edited

Reference Issue

Supersedes [#20254](#).

Context

Reductions over pain
filtering, cumulative s
many algorithms with

Various strategies exi
some using chunks (:

Merged

lorentzenchr merged 12 commits into [main](#) from [pairwise-distances-argkmin](#) on Feb 17, 2022

Conversation 17

Commits 12

Checks 0

Files changed 17

+2,063 -61



jjerphan commented on Jan 6, 2022

edited

Supersedes [#21462](#).

Relates to [#22587](#).

Task list as mentioned in [#21462 \(comment\)](#):

- Address the last reviews here as mentioned in [ENH Pairwise Distances ArgKmin #21462 \(comment\)](#)
- Finalize the sub-PR to use a class method [ENH Rework PairwiseDistancesArgKmin to use a class method .compute](#) [jjerphan/scikit-learn#5](#) for the small refactoring suggested by [@thomasjpfan](#).
- Refactor heap routines [MAINT refactor heap routines into utils #21987](#).

Reviewers



thomasjpfan



lorentzenchr



Assignees

No one—assign yourself

Labels

[cython](#) [module:metrics](#) [module:utils](#)

No Changelog Needed

Projects

None yet

101 of Community Building

coordination and decision making

many hard decisions

- new features require considerations on maintainability, stability, usability, cross-project collaboration (interoperability, long-running regimes)
- finding most stable and least intrusive way to solve a problem

"Think of a contribution like a puppy: you might view it as this cute, wonderful thing you're giving me while I'm looking at it as over a decade of feeding, walking, and vet bills."

Blog by Python developer Brett Cannon "Setting expectations for open source participation"

<https://snarky.ca/setting-expectations-for-open-source-participation/#submittingacontribution>

101 of Community Building

coordination and decision making

“The Hard Parts of Open Source” talk by Elm founder Evan Czaplicki

Sept 27-28, 2018
thewstrangeloop.com

https://www.youtube.com/watch?v=o_4EX4dPppA



101 of Community Building

coordination and decision making

“The Hard Parts of Open Source” talk by Elm founder Evan Czaplicki

Pattern

“All discussion is constructive”

Should Elm have user-definable infix operators?

“Yes, because there are cases where it makes code shorter and more convenient.”

“No, because it can make code harder to read, especially in a large team.”

flexibility ↑
↓ simplicity

Priorities

Constructive discussion seeks mutual understanding, rather than mutual agreement.

Sept 27-28, 2018
thestrangeloop.com

https://www.youtube.com/watch?v=o_4EX4dPppA



101 of Community Building

coordination and decision making

Nothing is ever set in stone

Scikit-learn enhancement proposals

latest ▾

Search docs

ACCEPTED

- SLEP000: SLEP and its workflow
- SLEP006: Metadata Routing
- SLEP007: Feature names, their generation and the API
- SLEP009: Keyword-only arguments
- SLEP010: `n_features_in_` attribute
- SLEP017: Clone Override Protocol with `__sklearn_clone__`
- SLEP018: Pandas Output for Transformers with `set_output`
- SLEP020: Simplifying Governance Changes

UNDER REVIEW

- SLEP017: Clone Override Protocol with `__sklearn_clone__`

Scikit-learn enhancement proposals

View page source

Scikit-learn enhancement proposals

This repository is for structured discussions about large modifications or additions to scikit-learn.

The discussions must create an "enhancement proposal", similar Python enhancement proposal, that reflects the major arguments to keep in mind, the rational and usecases that are addressed, the problems and the major possible solution. It should be a summary of the key points that drive the decision, and ideally converge to a draft of an API or object to be implemented in scikit-learn.

The SLEPs are publicly available online on [Read The Docs](#).

Accepted

- SLEP000: SLEP and its workflow
- SLEP006: Metadata Routing
- SLEP007: Feature names, their generation and the API
- SLEP009: Keyword-only arguments
- SLEP010: `n_features_in_` attribute
- SLEP017: Clone Override Protocol with `__sklearn_clone__`
- SLEP018: Pandas Output for Transformers with `set_output`



101 of Community Building

defining roles

Contributor Experience Team

- assist with triaging issues and PRs
- improve the overall experience for contributors

Documentation Team

- improve and review documentation of the project

Maintainers Team

- continued development of the project through ongoing engagement with the community

Communication Team

- outreach and communication for scikit-learn

101 of Community Building

mentoring

mentoring people on grants



CHAN
ZUCKERBERG
INITIATIVE



101 of Community Building

mentoring

dozens of community sprints via PyLadies and on Python conferences

The Value of Open Source Sprints, the scikit-learn Experience

2022-07-13 • 16 minute read

The blog post cover features a yellow-to-green gradient background with the scikit-learn logo. The title 'The Value of Open Source Sprints, the scikit-learn Experience' is at the top. Below it is a list of bullet points: '• 12 years', '• 45 sprints', '• 2400+ contributors', and a bell icon followed by 'What we have learned & are sharing'. At the bottom is a blue button with white text: 'blog.scikit-learn.org'.

Author:  [Reshma Shaikh](#)

<https://blog.scikit-learn.org/events/sprints-value/>

Probability 1.0

Not Just Code!

Building Sustainable OSS Communities



First scikit-learn sprint, Paris, March 2010; Photo credit: [Fabian Pedregosa](#); (from left to right): Fabian Pedregosa, Gael Varoquaux, Olivier Grisel, Alexandre Gramfort

101 of Community Building

mentoring

dozens of community sprints via PyLadies and on Python conferences

The Value of Open Source Sprints, the scikit-learn Experience

2022-07-13 • 16 minute read

The blog post cover features a yellow-to-green gradient background with the scikit-learn logo. The title 'The Value of Open Source Sprints, the scikit-learn Experience' is at the top. Below it is a list of bullet points: '• 12 years', '• 45 sprints', '• 2400+ contributors', and a bell icon followed by 'What we have learned & are sharing'. At the bottom left is a blue button with a white 'blog.' icon and the URL 'blog.scikit-learn.org'.

Author:  [Reshma Shaikh](#)

<https://blog.scikit-learn.org/events/sprints-value/>

Probability 1.0

Not Just Code! Building Sustainable OSS Communities



101 of Community Building

mentoring

mentoring outside of structures is hard

“The easiest way would be to pick from the issues, but sometimes while you are still thinking about a feasible solution, someone else may have already taken that issue. Therefore, we must have the ability to quickly figure out what features are “contributable”, as well as the courage to take some issues even if we are not 100% percent sure that we can come up with some quick solution.”

“Though we are able to do the maintenance issues (validating public functions) as mentioned in my previous posts, most of our past pull request are not being reviewed, and there are not many new issues to work on.”

<https://ossd-s23.github.io/Charlie-XIAO-weekly/>

101 of Community Building

communication between strangers

- mismatch in expectations
 - new contributors search in vain for ways to contribute
 - almost all tasks require high skills and a lot of contextual knowledge
 - results in new contributors sometimes guessing a solution and heavily relying on maintainers for correction and guidance



101 of Community Building

communication between strangers

expectation management

Finish Python 3.10 -> 3.11 bump #32650

 Closed  #32646

lesteve opened 2 weeks ago · edited by lesteve

Edits Member ...

 **Warning**

This issue is already being worked on, a comment like "can I work on this" or "assign this issue to me" may get you banned from the repo

From git grep below

- 3.10 -> 3.11 README.rst, pyproject.toml, meson.build, doc/developers/development_setup.rst ...
- update_environment_and_lock_files.py the version is misleading, it should be 3.12.3 in Ubuntu Atlas. For completeness, this is the Python version to create the pip requirements lock-file. ...
- Some possible clean-up in sklearn/preprocessing/_encoders.py ...
- .circleci/config.yml probably not used much since we are using the conda env, either bump or remove ...
- maybe other things I have missed ...



101 of Community Building

communication between strangers

expectation management



Add links to examples from the docstrings and user guide #30621

Expectation management

Helping users find the right information among our 10.000 pages of documentation is a complex and ongoing effort. If this task was trivial or fully mapped out, we'd have finished it already. If it was solvable by using an ILM-based tool, we'd have finished it already. But fact is: we need your support, your thoughtfulness and your critical thinking skills.

⚠ Important: we expect that only about 50% of the listed examples will ultimately be linked. Your contribution includes deciding whether an example adds enough value to be referenced, and if so, where. We are aware that this is not easy, especially for new contributors. We encourage you to share your reasoning, and a team member will make the final call. Even if your example isn't linked, your evaluation is still valuable.

By working on this issue, you share responsibility with us for creating good documentation for millions of users. That means thoughtful contributions matter. Please take the time to understand the context and consider your changes carefully before opening a PR. We cannot afford to accept low-effort contributions and will close PRs that do not follow the guidelines outlined in this issue.

How long will your first PR take you up until the point you open a PR?

- 8-16 hours if you have never contributed to a project and have only basic or no understanding of the workflow yet
- 2-8 hours if you're familiar with the general workflow but new to scikit-learn (closer to 2 hours if you're comfortable with linting, Sphinx, and CI)
- 1-2 hours for your 2nd, 3rd, ... PR on the same issue for everyone
- If it takes less time than this, your PR likely needs significant work after submission; and we want to avoid that.

How long will it take us to merge your PR?

- we strive for a scikit-learn member to look at your PR within a week and suggest changes depending on technical quality of the PR and an assessment of added value to the user by having that additional link in the documentation
- we strive for a maintainer to evaluate your PR within a month; they might also suggest changes before approving and merging
- the whole process on average takes several weeks and can take up to months, depending of availability of maintainers and on how many review cycles are necessary



101 of Community Building

communication between strangers

expectation management

This project is a community effort, and everyone is welcome to contribute. It is hosted on <https://github.com/scikit-learn/scikit-learn>.

▼ Comment on lines R27 to R28



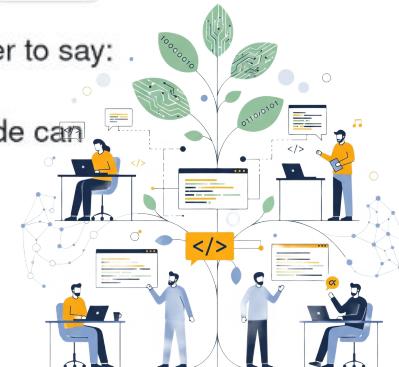
reshamas 2 days ago

Member

I think the "and everyone is welcome to contribute" is a bit confusing. Because really it would be better to say: .

scikit-learn is over 15 years old, and it is a mature open source project. Contributing to the project code can generally require advanced skills, and it may not be the best place to begin for new contributors.

etc.....



101 of Community Building

communication between strangers

expectation management

DOC Set expectations for contributors proposing significant work

Merged lesteve merged 5 commits into scikit-learn:main from betatim:maintainer-ingredient last week

Conversation 37 Commits 5 Checks 33 Files changed 1

 betatim commented 2 weeks ago • edited Member ...

A maintainer who has the time and interest is an important ingredient in moving work forward. The goal of this PR is to make it easier to make people aware of this.

The idea is to provide a label ("Needs Decision") that can be applied to an issue that will trigger a comment to be added that explains that, based on past experience, work that has no maintainer to champion it is unlikely to be merged. The goal is to raise awareness about this fact to help set expectations. The goal is not to use this to discourage people or suggest that an issue is not interesting/important/useful. The goal is to explain why sometimes even very good ideas don't happen for a long time.

Why community building is all worth it

- premise for an OSS project to exist
- feedback
 - relevance of features (use cases)
 - project stability (bug reports)
 - future developments (feature requests)
- secures future workforce
- innovation
 - voice needs and critique
 - fresh ideas

"It's more fun!"

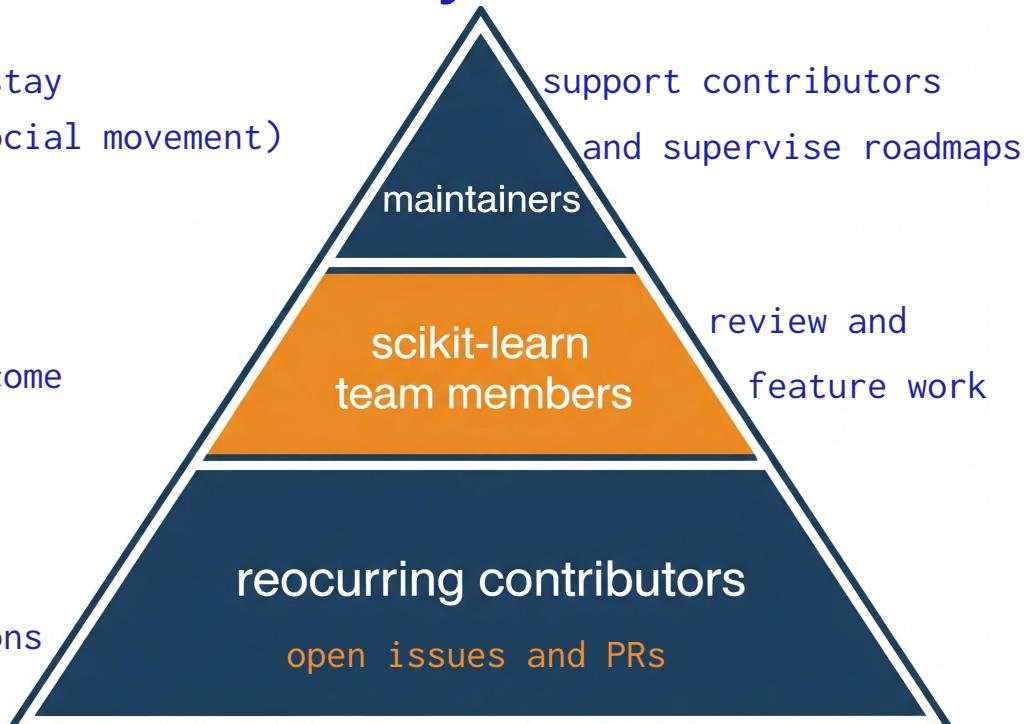


Anatomy of a healthy OSS community

- need for supply of new people who stay
- people come for values (OSS is a social movement)

individuals are looking for:

- healthy atmosphere and feeling welcome
- impact / making a difference
- recognition of their efforts
- mentorship
- inclusion / being heard for decisions
(depending on involvement)



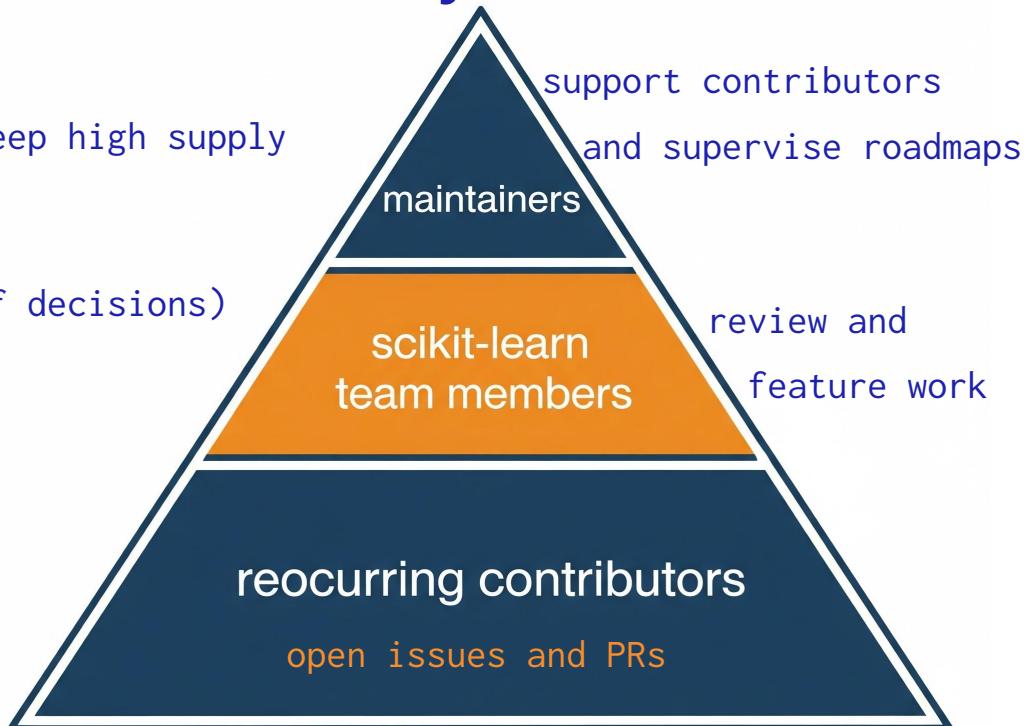
Anatomy of a healthy OSS community

community should care about:

- inclusivity, low entry threshold to keep high supply in potential new members
- mentoring pathways
- clarity in governance (transparency of decisions)
- flexibility and renewability

indicators of community health:

- low response latency
- diversity of contribution types
- newcomers return and grow
- progression of contributors becoming maintainers



OSS community and enterprise

symbiotic relationship

benefits for open source projects:

- monetary
 - sponsor contributors
 - project based partnerships
- infrastructure
 - SaaS, hardware, trainings, venues for sprints and summits, outreach and marketing



benefits for enterprises:

- features users actually need
- innovation lab
- funnel for finding talent
- strategic partnership

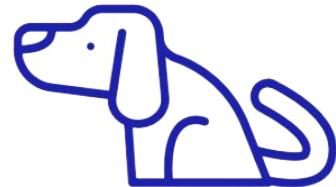
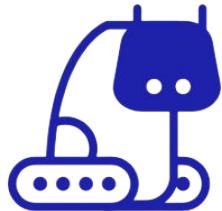
OSS community and enterprise

symbiotic relationship



Company involvement can strengthen an OSS community if they help to keep OSS communities healthy and uphold boundaries on preserving open, community-driven governance.

Discussions



Happy to Connect With You



<https://www.linkedin.com/in/StefanieSenger>

@StefanieSenger

Probability 1.0

