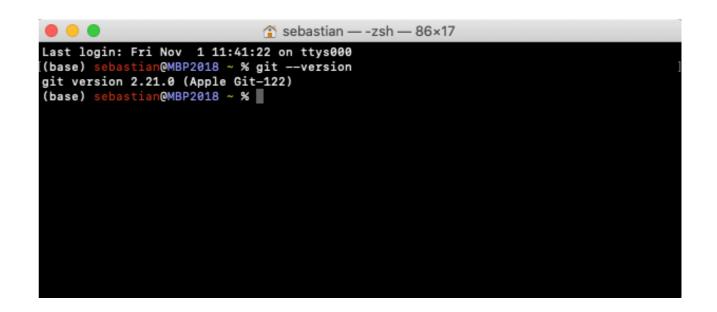
Contributing to Open Source Projects

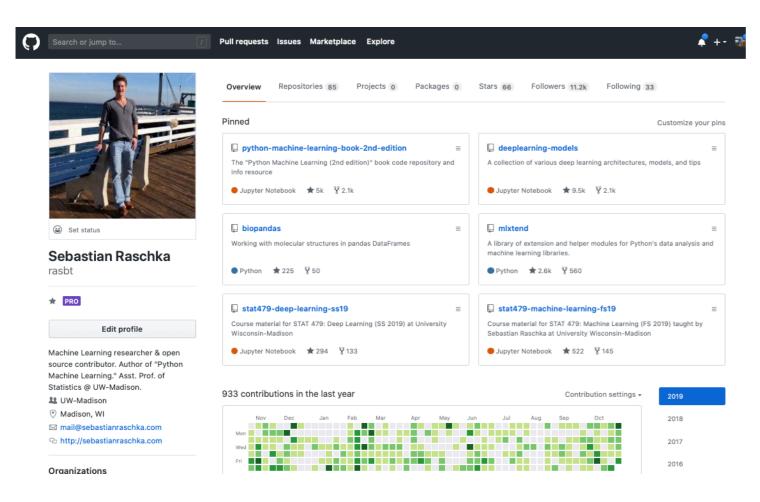
(like scikit-learn, mlxtend, etc.)

First Step: Familiarize yourself with Git & GitHub

Git vs. GitHub

Git is a software tool

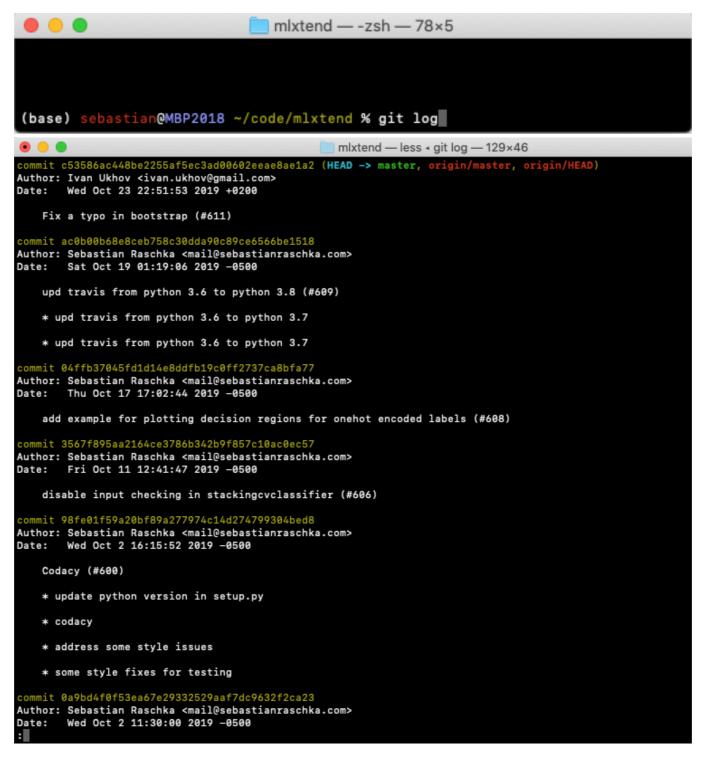




GitHub is a (web) platform for hosting "Git" projects

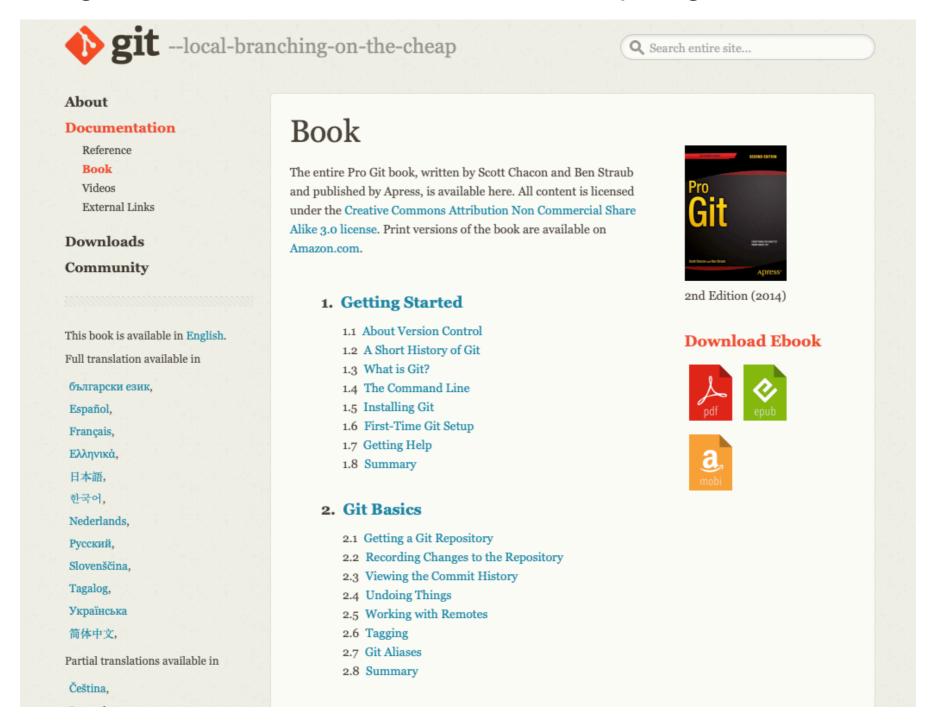
Git

- A version control system for software development
- Let's you keep track of changes in your code base





- Moreover, it's great for collaborative code development
- I.e., it's great for managing software development that involves multiple people
- A great learning resource is the free "Git" book at https://git-scm.com/book/en/v2



Additional Git Resources

- 1) What is Git | What is GitHub | Git Tutorial | GitHub Tutorial | Devops Tutorial | Edureka https://www.youtube.com/watch?v=xuB1Id2Wxak
- 2) Understanding branches in Git https://blog.thoughtram.io/git/rebase-book/2015/02/10/understanding-branches-in-git.html
- 3) .1 Git Branching What a Branch Is https://git-scm.com/book/en/v1/Git-Branching-What-a-Branch-Is
- 4) Forking Workflow https://www.atlassian.com/git/tutorials/comparing-workflows/forking-workflow

Basic Workflow for (most) Open Source Projects

Step 0: Read the Contributor Guidelines

http://rasbt.github.io/mlxtend/CONTRIBUTING/

mixtend Home User Guide → API → Installation About → Q Search C GitHub

How to Contribute

Quick Contributor Checklist

Tips for Contributors

Getting Started - Creating a New Issue and Forking the Repository

Syncing an Existing Fork

*The Main Workflow - Making Changes in a New Topic Branch

Notes for Developers

Building the documentation

Uploading a new version to PyPI

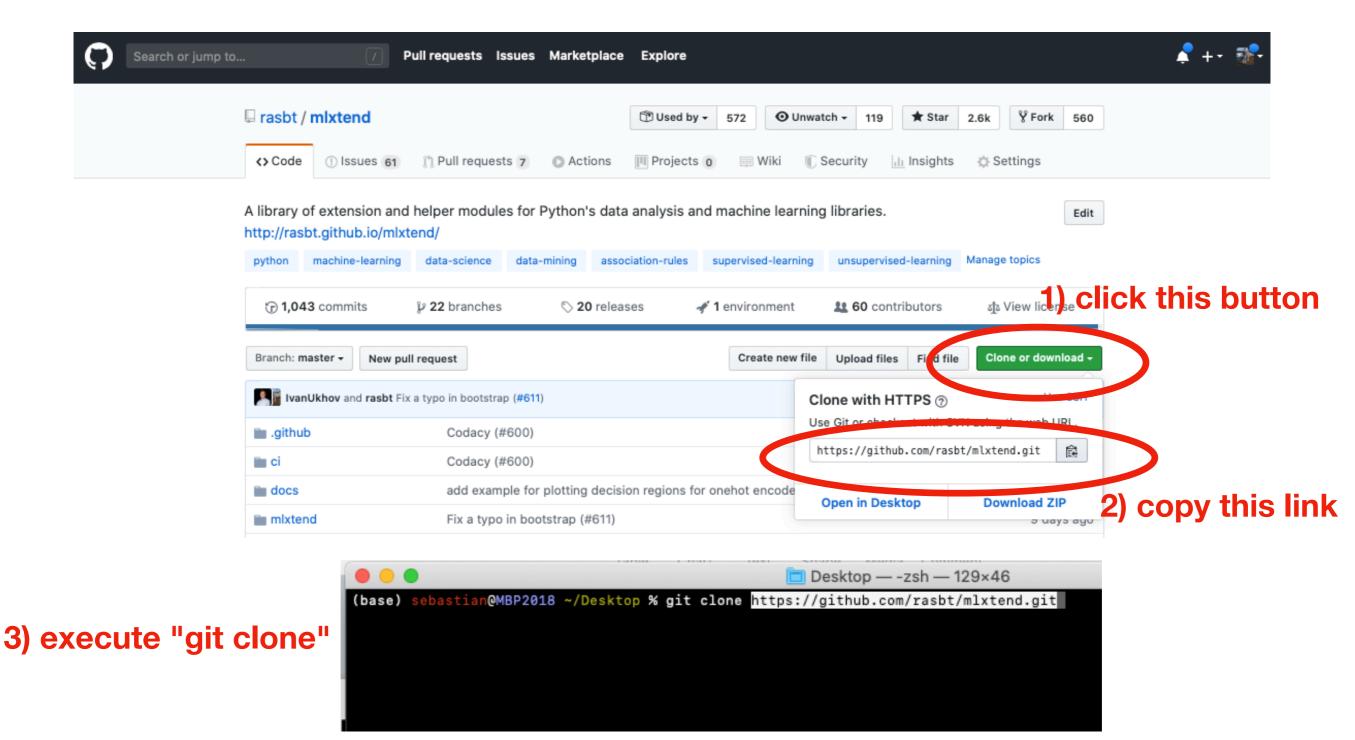
How to Contribute

I would be very happy about any kind of contributions that help to improve and extend the functionality of mlxtend.

Quick Contributor Checklist

This is a quick checklist about the different steps of a typical contribution to mlxtend (and other open source projects). Consider copying this list to a local text file (or the issue

Basic Workflow for (most) Open Source Projects Step 1: Clone the repository



Sebastian Raschka

STAT 479: Machine Learning

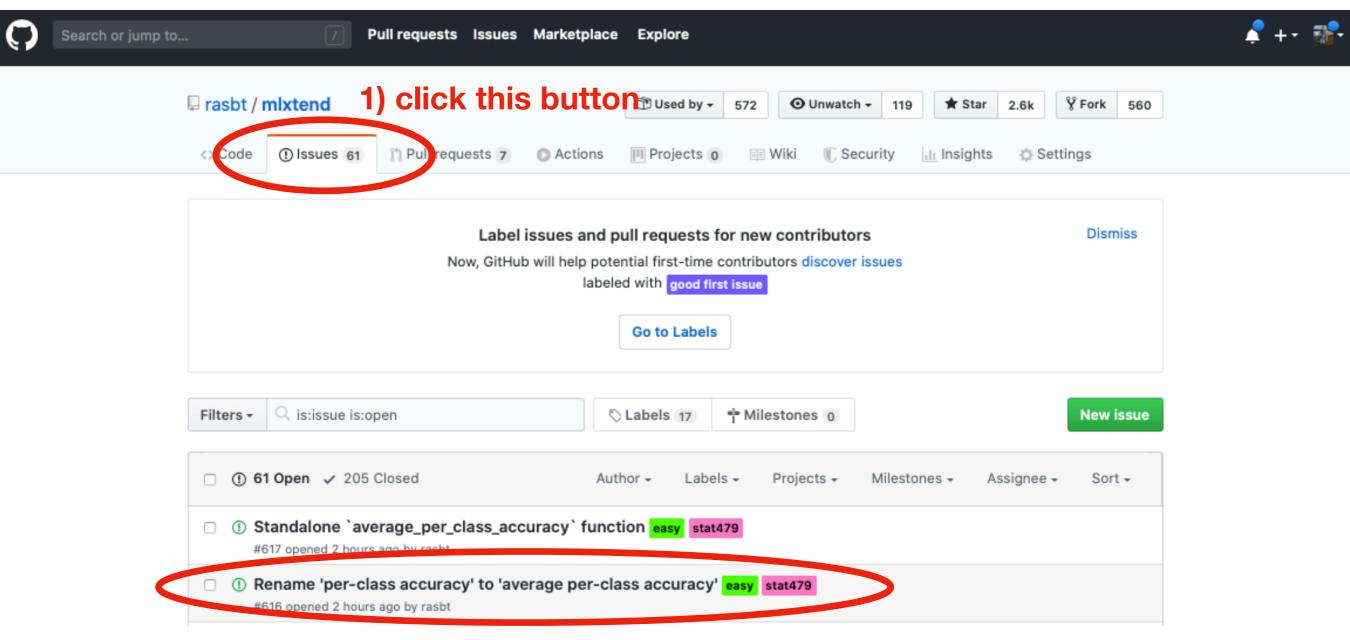
FS 2019

Basic Workflow for (most) Open Source Projects Step 1: Clone the repository

4) Navigate to the cloned directory

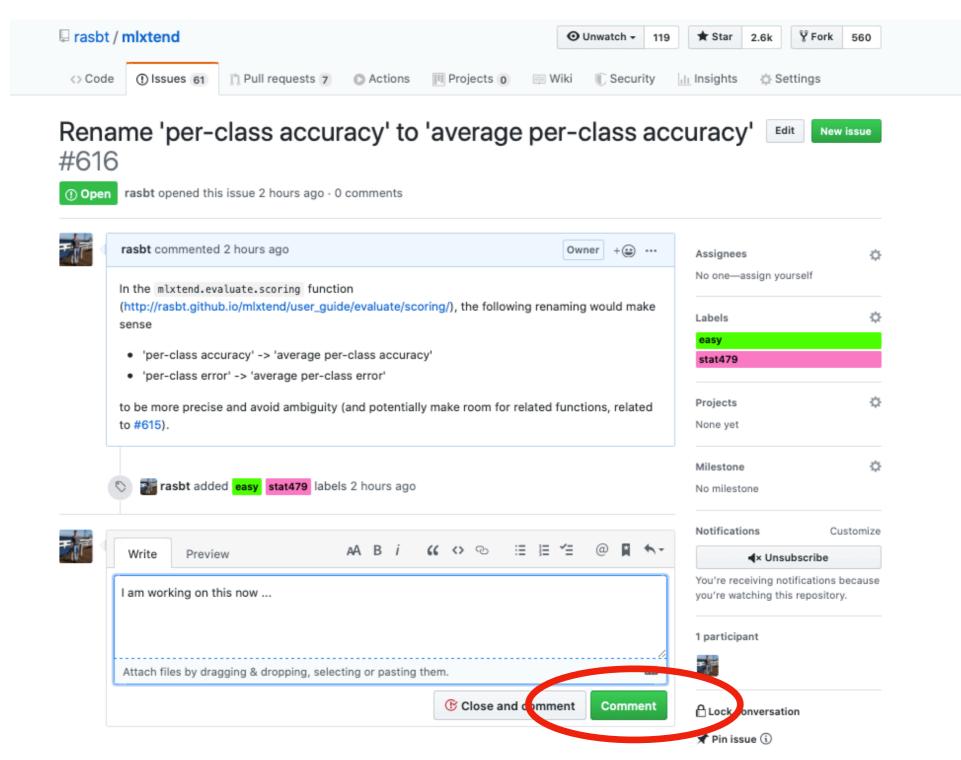
```
[(base) sebastian@MBP2018 ~/Desktop % git clone https://github.com/rasbt/mlxtend.git Cloning into 'mlxtend'...
remote: Enumerating objects: 71, done.
remote: Counting objects: 100% (71/71), done.
remote: Compressing objects: 100% (59/59), done.
remote: Total 26054 (delta 27), reused 35 (delta 12), pack-reused 25983
Receiving objects: 100% (26054/26054), 67.61 MiB | 2.30 MiB/s, done.
Resolving deltas: 100% (15167/15167), done.
[(base) sebastian@MBP2018 ~/Desktop % cd mlxtend
[(base) sebastian@MBP2018 ~/Desktop/mlxtend % git branch
* master
(base) sebastian@MBP2018 ~/Desktop/mlxtend %
```

Step 2: Find an Issue to Work On



2) click on an issue that interests you

Step 2: Find an Issue to Work On



3) add a comment to discuss the plan with the maintainer before working on it

11

Step 3: Keep Everything in Sync

- It's a good idea to periodically check that everything in your cloned directory is in sync with the original project on GitHub.
- For this, you have to setup an "upstream" link to the original repository (you only have to do it once after you cloned the repository!)

```
mlxtend — -zsh — 129×46

[(base) sebastian@MBP2018 ~/Desktop/mlxtend % git remote add upstream https://github.com/rasbt/mlxtend.git
[(base) sebastian@MBP2018 ~/Desktop/mlxtend % git remote -v
origin https://github.com/rasbt/mlxtend.git (fetch)
origin https://github.com/rasbt/mlxtend.git (push)
upstream https://github.com/rasbt/mlxtend.git (fetch)
upstream https://github.com/rasbt/mlxtend.git (push)
(base) sebastian@MBP2018 ~/Desktop/mlxtend %
```

Step 3: Keep Everything in Sync

1) Before starting working on a new feature, fetch changes from the original repo

```
mlxtend — -zsh — 129×46

[(base) sebastian@MBP2018 ~/Desktop/mlxtend % git remote add upstream https://github.com/rasbt/mlxtend.git
[(base) sebastian@MBP2018 ~/Desktop/mlxtend % git remote -v
origin https://github.com/rasbt/mlxtend.git (fetch)
origin https://github.com/rasbt/mlxtend.git (push)
upstream https://github.com/rasbt/mlxtend.git (fetch)
upstream https://github.com/rasbt/mlxtend.git (push)
(base) sebastian@MBP2018 ~/Desktop/mlxtend % git fetch upstream
```

2) Then merge the changes with the master branch

```
mlxtend — -zsh — 129×46
                n@MBP2018 ~/Desktop/mlxtend % git remote add upstream https://github.com
(base)
               n@MBP2018 ~/Desktop/mlxtend % git remote -v
(base)
origin https://github.com/rasbt/mlxtend.git (fetch)
origin https://github.com/rasbt/mlxtend.git (push)
                https://github.com/rasbt/mlxtend.git (fetch)
upstream
                https://github.com/rasbt/mlxtend.git (push)
upstream
             ian@MBP2018 ~/Desktop/mlxtend % git fetch upstream
(base)
From https://github.com/rasbt/mlxtend
* [new branch]
                       adaline-docs
                                             -> upstream/adaline-docs
                       categorical-decision -> upstream/categorical-decision
 * [new branch]
                                             -> upstream/codacy
 * [new branch]
                       codacy
  [new branch]
                       docfix
                                              -> upstream/docfix
                                             -> upstream/drop-proba
 * [new branch]
                       drop-proba
  [new branch]
                       fix-feat-imp
                                             -> upstream/fix-feat-imp
 * [new branch]
                       fixed-features
                                             -> upstream/fixed-features
   [new branch]
                       gh-pages
                                              -> upstream/gh-pages
 * [new branch]
                       grid-support-feat-sele -> upstream/grid-support-feat-sele
   [new branch]
                                             -> upstream/heatmap
                       heatmap
 * [new branch]
                       ldaloadings
                                             -> upstream/ldaloadings
   [new branch]
                                              -> upstream/master
                       master
 * [new branch]
                                             -> upstream/mnist
                       mnist
  [new branch]
                       pca-ddof
                                              -> upstream/pca-ddof
 * [new branch]
                       pca-svd
                                             -> upstream/pca-svd
   [new branch]
                                              -> upstream/py37
 * [new branch]
                       readme-links
                                             -> upstream/readme-links
 * [new branch]
                       sclf-warn
                                              -> upstream/sclf-warn
 * [new branch]
                       setup
                                             -> upstream/setup
  [new branch]
                       stacking-dtype
                                              -> upstream/stacking-dtype
 * [new branch]
                                             -> upstream/v0.17.0
                       v0.17.0
* [new branch]
                       verbose-apriori
                                              -> upstream/verbose-apriori
                n@MBP2018 ~/Desktop/mlxtend % git merge upstream/master
(base)
Already up to date.
           stian@MBP2018 ~/Desktop/mlxtend %
```

Step 4: Make a New Feature Branch

- Now, create a new feature branch, which will be the branch where you will work
 on the new feature you are planning to implement or issue you are going to fix
- Then, execute "git branch" to ensure that you are on the correct branch now

```
mlxtend — -zsh — 129×46

[(base) sebastian@MBP2018 ~/Desktop/mlxtend % git checkout -b rename-per-class-accuracy
Switched to a new branch 'rename-per-class-accuracy'
[(base) sebastian@MBP2018 ~/Desktop/mlxtend % git branch
master

* rename-per-class-accuracy
```

Step 5: Start Coding

You can start working on the feature/issue now

```
■ Implementation of the model of the mod
```

```
scoring.py
       🕏 scoring.py 🗨
       Users > sebastian > Desktop > mlxtend > mlxtend > evaluate > 🕏 scoring.py > 😚 scoring
        22
                  scores = []
        23 🗸
                  for l in unique labels:
                      scores.append(func(np.where(true != l, 1, 0),
        24 🗸
        25
                                          np.where(pred != l, 1, 0)))
                  return float(sum(scores)) / len(scores)
        26
        27
宓
        28
        29 vdef scoring(y_target, y_predicted, metric='error',
                          positive_label=1, unique_labels='auto'):
        30
留
                  """Compute a scoring metric for supervised learning.
        31
        32
        33
                  Parameters
        34
                  y_target : array-like, shape=[n_values]
        35 🗸
                      True class labels or target values.
        36
                  y_predicted : array-like, shape=[n_values]
        37 🗸
                      Predicted class labels or target values.
        38
                  metric : str (default: 'error')
        39 🗸
        40
                      Performance metric:
                      'accuracy': (TP + TN)/(FP + FN + TP + TN) = 1-ERR\n
        41
        42
                      'average per-class accuracy': Average per-class accuracy\n
                      'average per-class error': Average per-class error\n
```

Step 6: Code Testing

After developing the new code, also add or modify the test functions

```
mlxtend — -zsh — 129×46

(base) sebastian@MBP2018 ~/Desktop/mlxtend % code mlxtend/evaluate/tests/test_scoring.py
```

```
test_scoring.py
                       test_scoring.py
      scoring.py
      Users > sebastian > Desktop > mlxtend > mlxtend > evaluate > tests > 💠 test_scoring.py > 😚 test_avg_perclass_error
             def test_f1():
                 y_targ = [1, 1, 1, 0, 0, 1, 0, 1]
                 y_pred = [1, 0, 1, 0, 0, 0, 1, 1]
                 res = scoring(y_target=y_targ, y_predicted=y_pred, metric='f1')
                 assert round(res, 3) == 0.667, res
嵏
             def test_matthews_corr_coef():
                 y_targ = [1, 1, 1, 0, 0, 1, 0, 1]
                 y_pred = [1, 0, 1, 0, 0, 0, 1, 1]
                 res = scoring(y_target=y_targ,
                                y_predicted=y_pred,
                                metric='matthews_corr_coef')
                 assert round(res, 3) == 0.258, res
             def test_avg_perclass_accuracy():
               y_targ = np.array([0, 0, 0, 1, 1, 1, 1, 1, 2, 2])
                 y_pred = np.array([0, 1, 1, 0, 1, 1, 2, 2, 2, 2])
                  res = scoring(y_target=y_targ,
                                y_predicted=y_pred,
       123
                                metric='average per-class accuracy')
                  assert round(res, 3) == 0.667, res
             def test_avg_perclass_error():
                y_targ = np.array([0, 0, 0, 1, 1, 1, 1, 1, 2, 2])
                 y_pred = np.array([0, 1, 1, 0, 1, 1, 2, 2, 2, 2])
                 res = scoring(y_target=y_targ,
                                y_predicted=y_pred,
       132
                               metric='average per-class error')
                  assert round(res, 3) == 0.333, res
```

Step 6: Code Testing

 Then, test your code locally to make sure that your changes didn't break the existing code base

```
mlxtend — -zsh — 129×46
(base) sebastian@MBP2018 ~/Desktop/mlxtend % PYTHONPATH='.' pytest ./mlxtend -sv
```

```
mlxtend — -zsh — 129×46
mlxtend/regressor/tests/test_stacking_cv_regression.py::test_weight_unsupported_with_no_weight PASSED
mlxtend/regressor/tests/test_stacking_cv_regression.py::test_gridsearch_replace_mix PASSED
mlxtend/regressor/tests/test_stacking_regression.py::test_multivariate PASSED
mlxtend/regressor/tests/test_stacking_regression.py::test_multivariate_class PASSED
mlxtend/regressor/tests/test_stacking_regression.py::test_sample_weight PASSED
mlxtend/regressor/tests/test_stacking_regression.py::test_weight_ones PASSED
mlxtend/regressor/tests/test_stacking_regression.py::test_weight_unsupported_regressor PASSED
mlxtend/regressor/tests/test_stacking_regression.py::test_weight_unsupported_meta PASSED
mlxtend/regressor/tests/test_stacking_regression.py::test_weight_unsupported_with_no_weight PASSED
mlxtend/regressor/tests/test_stacking_regression.py::test_gridsearch PASSED
mlxtend/regressor/tests/test_stacking_regression.py::test_gridsearch_numerate_regr PASSED
mlxtend/regressor/tests/test_stacking_regression.py::test_get_coeff PASSED
mlxtend/regressor/tests/test_stacking_regression.py::test_get_intercept PASSED
mlxtend/regressor/tests/test_stacking_regression.py::test_get_coeff_fail PASSED
mlxtend/regressor/tests/test_stacking_regression.py::test_get_params PASSED
mlxtend/regressor/tests/test_stacking_regression.py::test_regressor_gridsearch PASSED
mlxtend/regressor/tests/test_stacking_regression.py::test_predict_meta_features PASSED
mlxtend/regressor/tests/test_stacking_regression.py::test_train_meta_features_ PASSED
mlxtend/regressor/tests/test_stacking_regression.py::test_not_fitted_predict PASSED
mlxtend/regressor/tests/test_stacking_regression.py::test_clone PAS
mlxtend/regressor/tests/test_stacking_regression.py::test_features_in_secondary 0.13732094333079276
0.12403094909404185
mlxtend/regressor/tests/test_stacking_regression.py::test_predictions_from_sparse_matrix 0.6082588436138667
mlxtend/regressor/tests/test_stacking_regression.py::test_sparse_matrix_inputs_and_features_in_secondary PASSED
mlxtend/text/tests/test_generalize_names.py::test_generalize_names PASSED
mlxtend/text/tests/test_generalize_names_duplcheck.py::test_generalize_names_duplcheck PASSED
mlxtend/text/tests/test_tokenizer.py::test_tokenizer_words_and_emoticons_1 PASSED
mlxtend/text/tests/test_tokenizer.py::test_tokenizer_words_and_emoticons_2 PASSED
mlxtend/utils/tests/test_checking_inputs.py::test_check_Xy_ok PASSED
mlxtend/utils/tests/test_checking_inputs.py::test_check_Xy_invalid_type_X PASSED
mlxtend/utils/tests/test_checking_inputs.py::test_check_Xy_float16_X PASSED
mlxtend/utils/tests/test_checking_inputs.py::test_check_Xy_float16_y PASSED
mlxtend/utils/tests/test_checking_inputs.py::test_check_Xy_invalid_type_y PASSED
mlxtend/utils/tests/test_checking_inputs.py::test_check_Xy_invalid_dtype_X PASSED
mlxtend/utils/tests/test_checking_inputs.py::test_check_Xy_invalid_dtype_y PASSED
```

Step 7: Update the Documentation

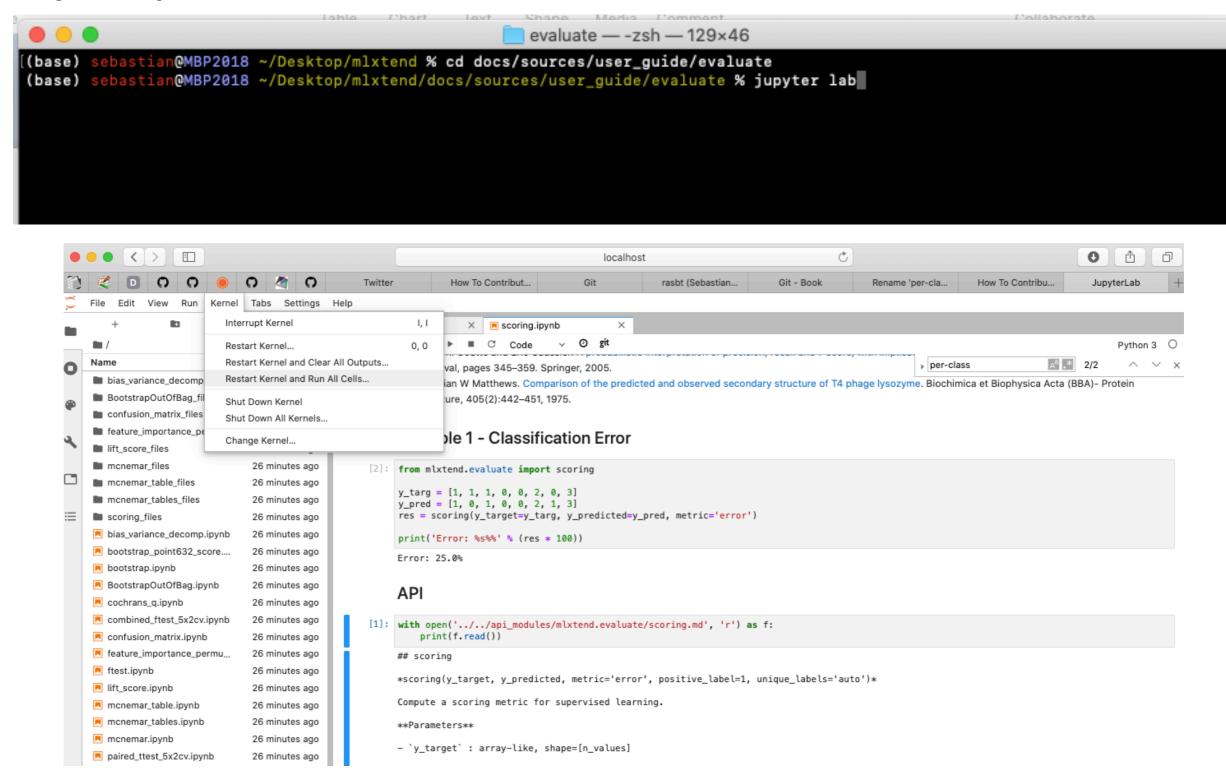
Run the python make_api.py file to update the general code documentation (this is mlxtend-specific)

```
docs — -zsh — 129×46

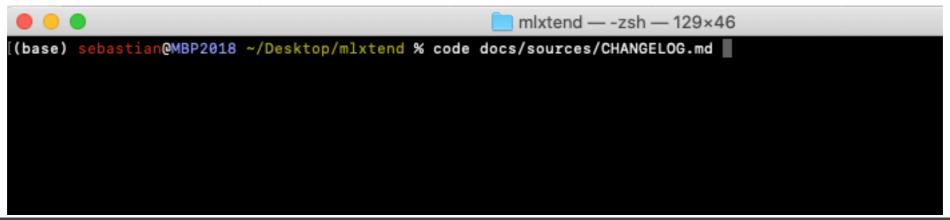
[(base) sebastian@MBP2018 ~/Desktop/mlxtend % cd docs
[(base) sebastian@MBP2018 ~/Desktop/mlxtend/docs % python make_api.py
```

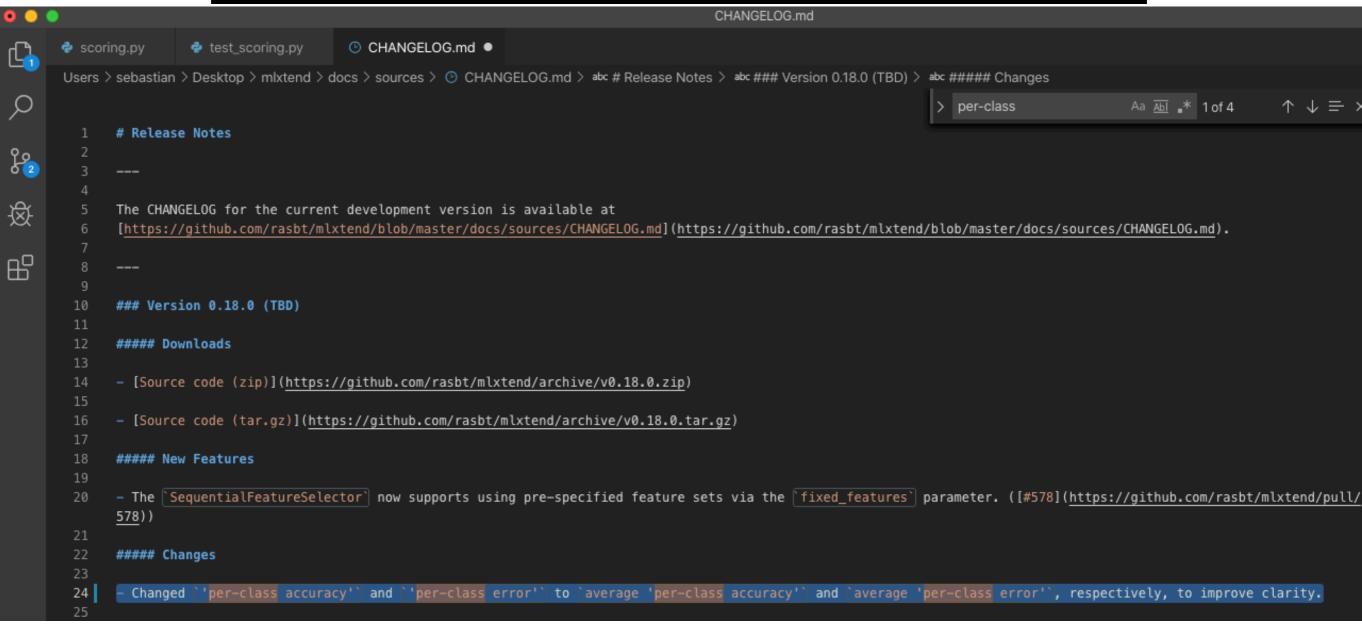
Step 7: Update the Documentation

 Next, update the documentation on the Jupyter notebook (this is mixtendspecific)



Step 8: Make a Changelog Entry





Use "git status" to see what files you have changes

```
[(base) sebastian@MBP2018 ~/Desktop/mlxtend % git status

On branch rename-per-class-accuracy

Changes not staged for commit:

(use "git add <file>..." to update what will be committed)

(use "git checkout -- <file>..." to discard changes in working directory)

modified: docs/sources/user_guide/evaluate/scoring.ipynb
modified: mlxtend/evaluate/scoring.py
modified: mlxtend/evaluate/tests/test_scoring.py

no changes added to commit (use "git add" and/or "git commit -a")

(base) sebastian@MBP2018 ~/Desktop/mlxtend %
```

- Then add the files you want to upload via "git add filename"
- If the "git status" list does not include any files that you don't want to upload, you can also execute "git add." to stage all files for uploading all files

```
mlxtend — -zsh — 129×46
[(base) sebastian@MBP2018 ~/Desktop/mlxtend % git status
On branch rename-per-class-accuracy
Changes not staged for commit:
  (use "git add <file>..." to update what will be committed)
  (use "git checkout -- <file>..." to discard changes in working directory)
                  docs/sources/user_guide/evaluate/scoring.ipynb
       modified: mlxtend/evaluate/scoring.py
       modified: mlxtend/evaluate/tests/test_scoring.py
no changes added to commit (use "git add" and/or "git commit -a")
(base) sebastian@MBP2018 ~/Desktop/mlxtend % git add .
(base) sebastian@MBP2018 ~/Desktop/mlxtend % git status
On branch rename-per-class-accuracy
Changes to be committed:
  (use "git reset HEAD <file>..." to unstage)
        modified: docs/sources/user_guide/evaluate/scoring.ipynb
       modified: mlxtend/evaluate/scoring.py
       modified: mlxtend/evaluate/tests/test_scoring.py
(base) sebastian@MBP2018 ~/Desktop/mlxtend %
```

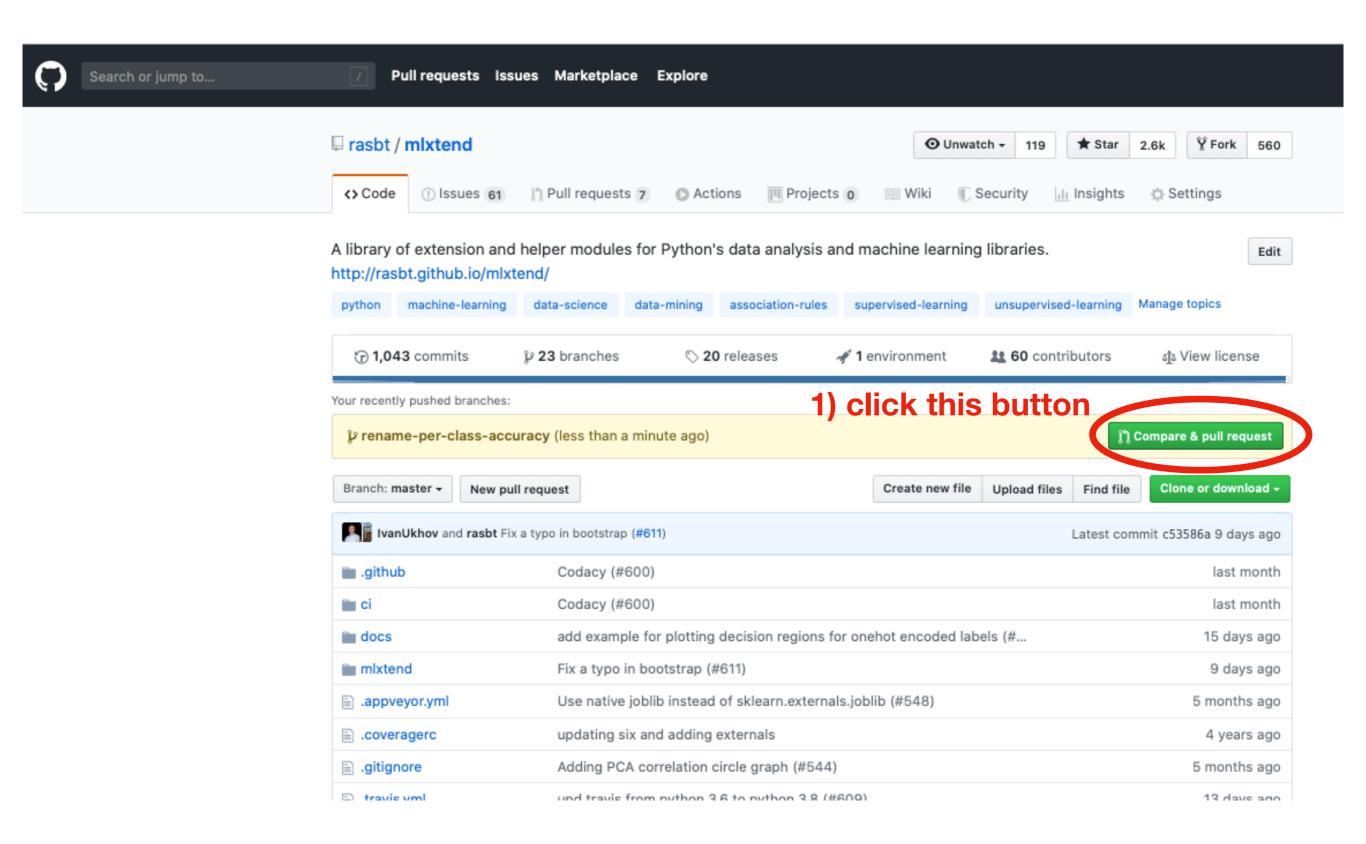
 After marking the files you want to upload via "git add", you can now use "git commit -m 'your message'" to make a entry in the GitHub history

```
. .
                                                     mlxtend — -zsh — 129×46
(base) sebastian@MBP2018 ~/Desktop/mlxtend % git status
On branch rename-per-class-accuracy
Changes not staged for commit:
  (use "git add <file>..." to update what will be committed)
  (use "git checkout -- <file>..." to discard changes in working directory)
       modified: docs/sources/user_guide/evaluate/scoring.ipynb
       modified: mlxtend/evaluate/scoring.py
                   mlxtend/evaluate/tests/test_scoring.py
no changes added to commit (use "git add" and/or "git commit -a")
(base) sebastian@MBP2018 ~/Desktop/mlxtend % git add .
(base) sebastian@MBP2018 ~/Desktop/mlxtend % git status
On branch rename-per-class-accuracy
Changes to be committed:
  (use "git reset HEAD <file>..." to unstage)
       modified: docs/sources/user_guide/evaluate/scoring.ipynb
       modified: mlxtend/evaluate/scoring.py
       modified: mlxtend/evaluate/tests/test_scoring.py
(base) sebastian@MBP2018 ~/Desktop/mlxtend % git commit -m 'rename "per-class accuracy" to "average per-class accuracy"'
```

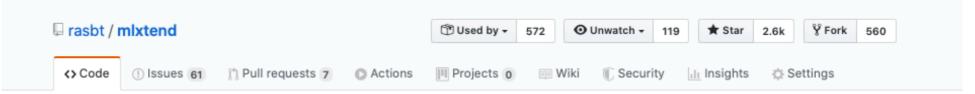
 Push your changes to the server

```
mlxtend — -zsh — 129×46
(base) sebastian@MBP2018 ~/Desktop/mlxtend % git status
On branch rename-per-class-accuracy
Changes not staged for commit:
  (use "git add <file>..." to update what will be committed)
  (use "git checkout -- <file>..." to discard changes in working directory)
                    docs/sources/user_guide/evaluate/scoring.ipynb
        modified: mlxtend/evaluate/tests/test_scoring.py
no changes added to commit (use "git add" and/or "git commit -a")
(base) sebastian@MBP2018 ~/Desktop/mlxtend % git add .
(base) sebastian@MBP2018 ~/Desktop/mlxtend % git status
On branch rename-per-class-accuracy
Changes to be committed:
  (use "git reset HEAD <file>..." to unstage)
        modified: docs/sources/user_guide/evaluate/scoring.ipynb
        modified: mlxtend/evaluate/scoring.py
        modified: mlxtend/evaluate/tests/test_scoring.py
(base) sebastian@MBP2018 ~/Desktop/mlxtend % git commit -m 'rename "per-class accuracy" to "average per-class accuracy"'
[rename-per-class-accuracy 182d81f1] rename "per-class accuracy" to "average per-class accuracy"
3 files changed, 14 insertions(+), 16 deletions(-)
(base) sebastian@MBP2018 ~/Desktop/mlxtend % git status
On branch rename-per-class-accuracy
nothing to commit, working tree clean
(base) sebastian@MBP2018 ~/Desktop/mlxtend % git branch
  master
rename-per-class-accuracy
(base) sebastian@MBP2018 ~/Desktop/mlxtend % git push origin rename-per-class-accuracy
 master
rename-per-class-accuracy
(base) sebastian@MBP2018 ~/Desktop/mlxtend % git push origin rename-per-class-accuracy
Enumerating objects: 23, done.
Counting objects: 100% (23/23), done.
Delta compression using up to 12 threads
Compressing objects: 100% (12/12), done.
Writing objects: 100% (12/12), 986 bytes | 986.00 KiB/s, done.
Total 12 (delta 11), reused 0 (delta 0)
remote: Resolving deltas: 100% (11/11), completed with 11 local objects.
remote:
remote: Create a pull request for 'rename-per-class-accuracy' on GitHub by visiting:
            https://github.com/rasbt/mlxtend/pull/new/rename-per-class-accuracy
remote:
To https://github.com/rasbt/mlxtend.git
 * [new branch]
                   rename-per-class-accuracy -> rename-per-class-accuracy
(base) sebastian@MBP2018 ~/Desktop/mlxtend %
```

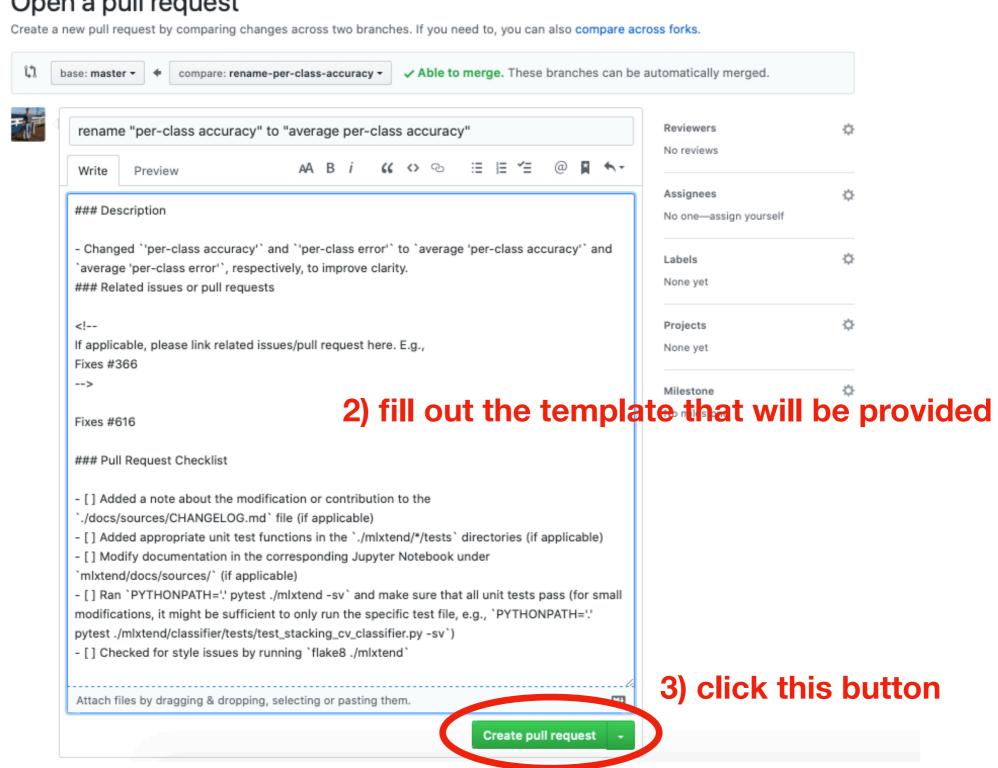
Step 10: Open a Pull Request



Step 10: Open a Pull Request

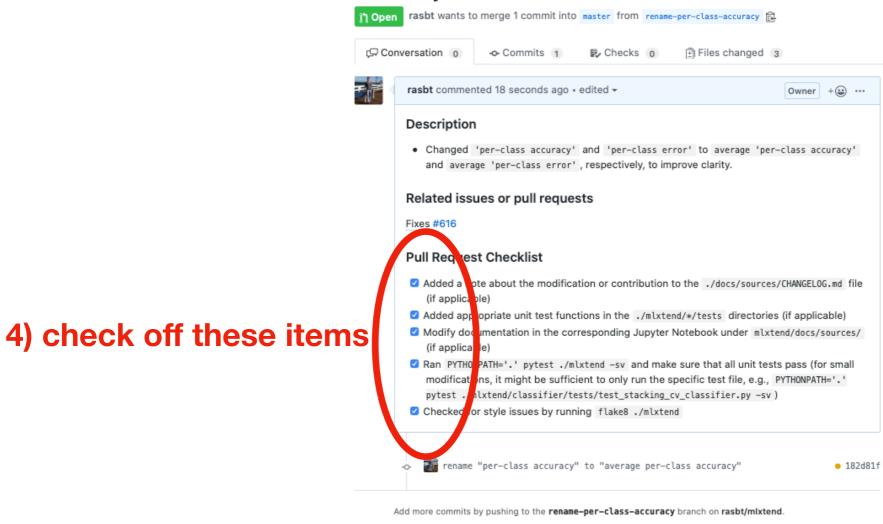


Open a pull request



Step 10: Open a Pull Request

rename "per-class accuracy" to "average per-class accuracy" #618



5) check the automated unit test results later

