



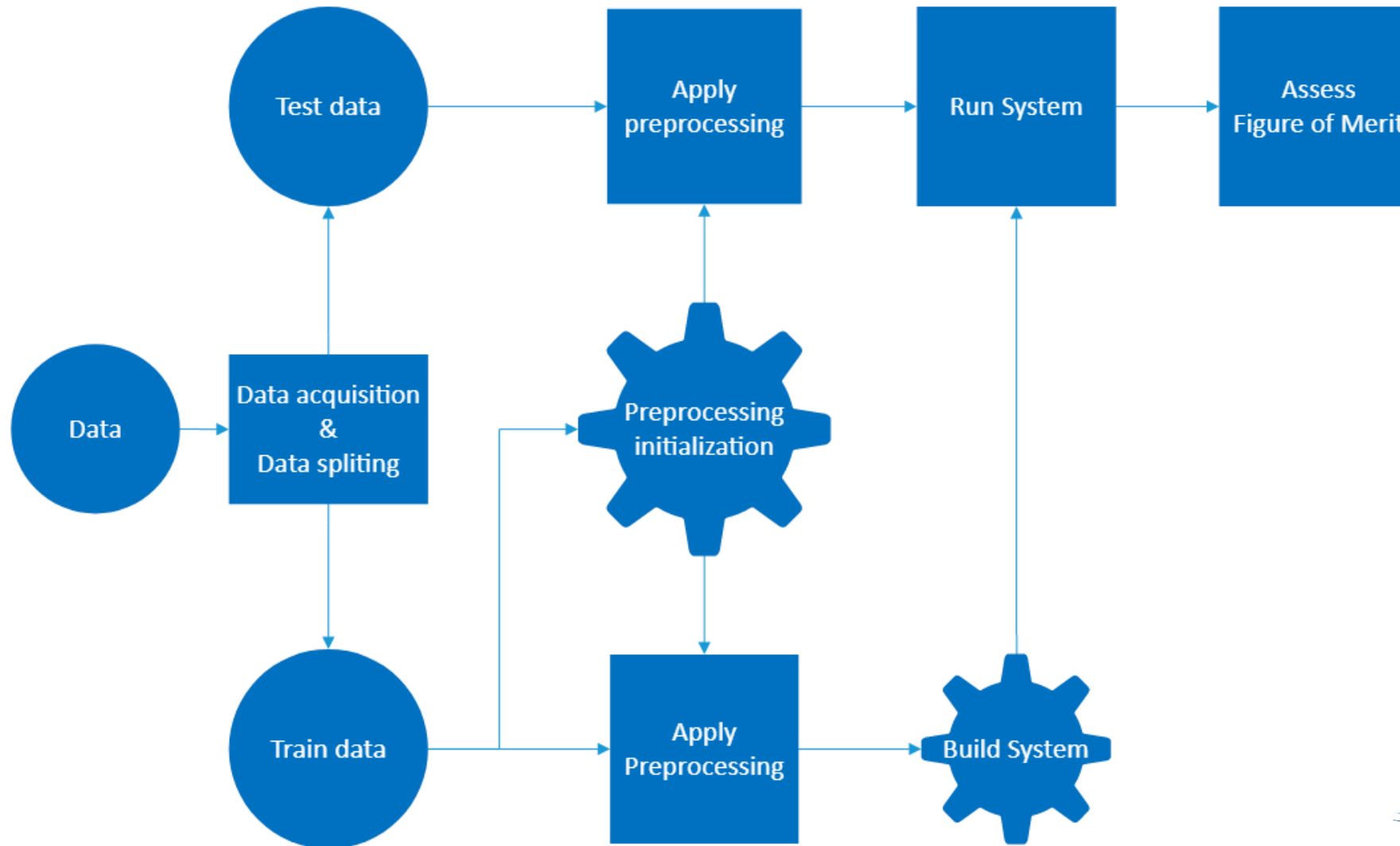
Mini-Project: Reproducibility in AI

Cédric Mariéthoz & Adrien Chassignet

Introduction

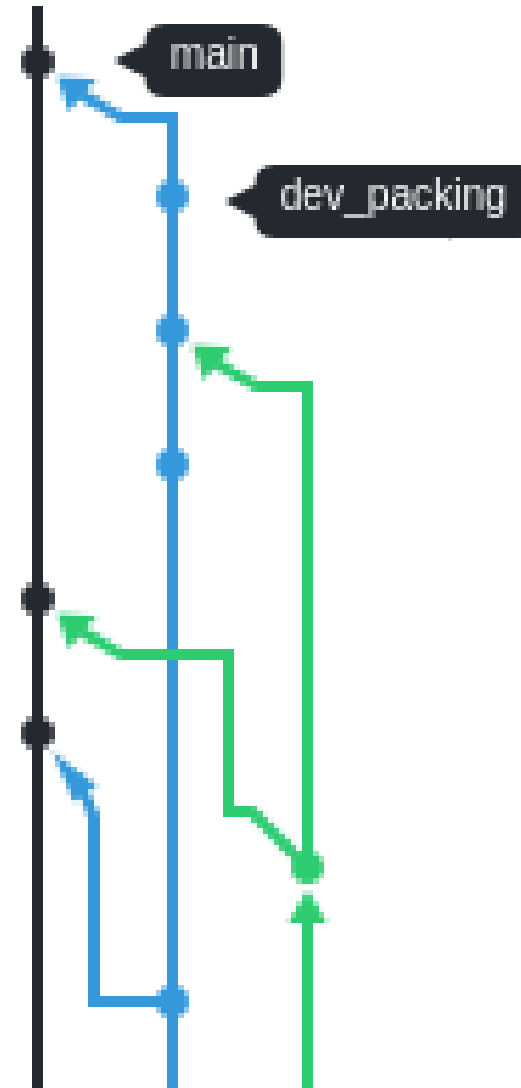
- Workflow
- Git
 - GitHub
- Unit Testing
 - Coverage
 - CI
- Documentation
- Packaging
- License

Data Workflow



git

- Basic command
 - add commit push pull
 - status log
- Branch
 - checkout merge branch
- Strategy
 - One branch per job
 - No work directly on main



GitHub

- Organization to host and share the mini-project repository
- Pull requests to add new features

Dev main module #6

Merged AdrienChassignet merged 3 commits into `main` from `dev-main-module` 9 days ago

Conversation 0 Commits 3 Checks 4 Files changed 3

AdrienChassignet commented 9 days ago

Refactored how command-line arguments were handled. Now the main module contains dictionaries for the different building blocks of the pipeline with keys as the name of the method and the value being the function to be called for the chosen block. This way the code is more extensible and if statements are avoided to parse the CL arguments.

AdrienChassignet added 3 commits 9 days ago

- Created `models.py` file and added custom classes inheriting from `sklea...` f567128
- Removed unnecessary and too restrictive assertions 2d2a2b5
- Refactored list of arguments to use dictionaries with functions to b... 4863539

AdrienChassignet merged commit `fd582df` into `main` 9 days ago

4 checks passed

AdrienChassignet deleted the `dev-main-module` branch 9 days ago

M05-project-group5/mini x +

https://github.com/M05-project-group5/mini-project

Search or jump to... Pull requests Issues Marketplace Explore

M05-project-group5 / mini-project Public

Pin Watch 0 Fork 0 Star 0

<> Code Issues 1 Pull requests Actions Projects Wiki Security Insights

main 4 branches 0 tags Go to file Add file Code

mariethoz Merge branch 'dev_test_preproces...' b478579 17 hours ago 79 commits

| | | |
|----------------------|---|--------------|
| .github/workflows | Update ci-testing.yml | 21 hours ago |
| src | Merge branch 'dev_preprocessing' into main | 18 hours ago |
| test | Merge branch 'dev_test_preprocessing' into main | 17 hours ago |
| .gitignore | Initial commit | 23 days ago |
| README.md | Update README.md | 20 hours ago |
| command.bat | Add test for preprocessing | 7 days ago |
| download_datasets.py | Added download functions specific to datasets | 14 days ago |
| environment.yml | Rename environment.yml to environment.yml | 21 hours ago |
| main.py | Add unit tests for the main module | 21 hours ago |

coverage 86%

Mini-project

This is an extensible and fully reproducible system to analyze multiple datasets, with various Machine Learning techniques.

Description

About

Extensible and fully reproducible system to analyze multiple datasets, with various Machine Learning techniques.

Readme

0 stars

0 watching

0 forks

Releases

No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)

Contributors 2

- AdrienChassignet
- mariethoz

Languages

Unit testing

- Using Pytest, a set of unit tests has been written for every module of the pipeline



Coverage

- Coveralls.io : keep track of the coverage and pin out missing tests

COMMITTED 21 MAR 2022 - 18:02

BUILD #

2017870703

BUILD TYPE

push

github

COMMITTED BY

AdrienChassinnet

COMMIT MESSAGE

Fix warning about deprecated operator when parsing Boston house prices dataset

RUN DETAILS

422 of 442 relevant lines covered (95.48%)

0.95 hits per line

JOBS

COVERAGE

↑ 95.48

JOB

3.8 - 2017870703.1

FILES COVERED

9

RAN

21 Mar 2022 06:02PM UTC

SOURCE FILES ON BUILD 2017870703

TREE

LIST 9

CHANGED 3

SOURCE CHANGED 2

COVERAGE CHANGED 3

SEARCH:

↑ COVERAGE

⇅ FILE

⇅ LINES

⇅ RELEVANT

⇅ COVERED

⇅ MISSED

⇅ HITS/LINE

↓ 77.94

main.py

130

68

53 -2

15 +2

1.0

↑ 85.19

src/load_data.py

95 -23

27 -15

23

4 -15

1.0

↓ 98.72

test/test_main.py

103

78

77

1

1.0

↻ 100.0

test/test_data_loading.py

50

37

37

0

1.0

↓ 100.0

test/test_preprocessing.py

170

120

120

0

1.0

↑ 100.0

download_datasets.py

74 +1

25 -5

25 +2

0 -7

1.0

↓ 100.0

test/test_split_data.py

89

53

53

0

1.0

↓ 100.0

src/preprocessing_data.py

70

25

25

0

1.0

↓ 100.0

src/split_data.py

43

9

9

0

1.0

SHOW 10

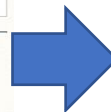
ENTRIES

Showing 1 to 9 of 9 entries

< PREVIOUS

1

NEXT >



COMMITTED 21 MAR 2022 - 18:47

COVERAGE INCREASED (+1.7%) TO 97.216%

BUILD #

2018100895

BUILD TYPE

push
github

COMMITTED BY

GitHub

COMMIT MESSAGE

Merge pull request #5 from M05-project-group5/dev-unit-test-main

RUN DETAILS

454 of 467 relevant lines covered (97.22%)
0.97 hits per line

JOBS

COVERAGE

JOB

FILES COVERED

RAN

↑ 97.22

3.8 - 2018100895.1

9

21 Mar 2022 06:47PM UTC

SOURCE FILES ON BUILD 2018100895

TREE

LIST 9

CHANGED 3

SOURCE CHANGED 1

COVERAGE CHANGED 3

SEARCH:

COVERAGE

FILE

LINES

RELEVANT

COVERED

MISSED

HITS/LINE

↑ 91.18

main.py

130

68

62 +9

6 -9

1.0

↓ 93.2

test/test_main.py

134 +31

103 +25

96 +19

7 +6

1.0

— 100.0

src/preprocessing_data.py

70

25

25

0

1.0

— 100.0

src/split_data.py

43

9

9

0

1.0

— 100.0

test/test_data_loading.py

50

37

37

0

1.0

— 100.0

test/test_preprocessing.py

170

120

120

0

1.0

— 100.0

download_datasets.py

74

25

25

0

1.0

— 100.0

test/test_split_data.py

89

53

53

0

1.0

↑ 100.0

src/load_data.py

95

27

27 +4

0 -4

1.0

SHOW 10 ENTRIES

Showing 1 to 9 of 9 entries

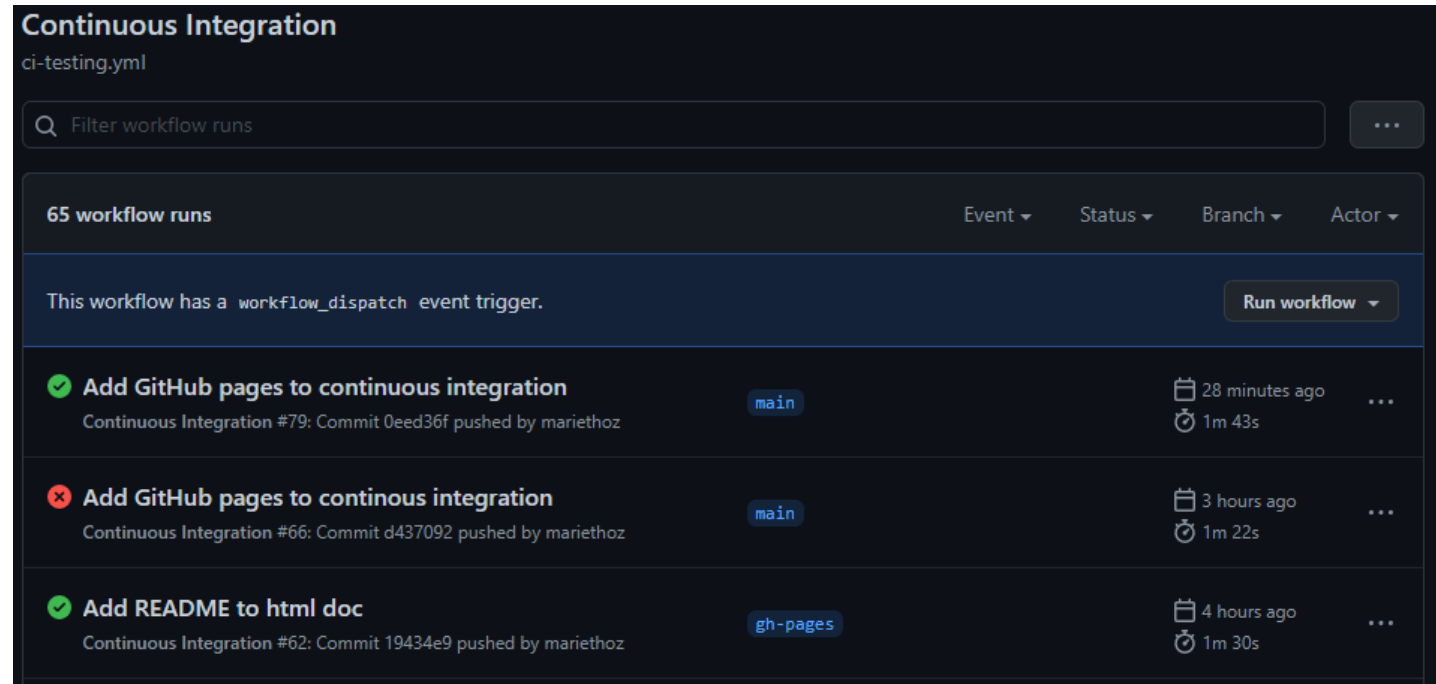
< PREVIOUS

1

NEXT >

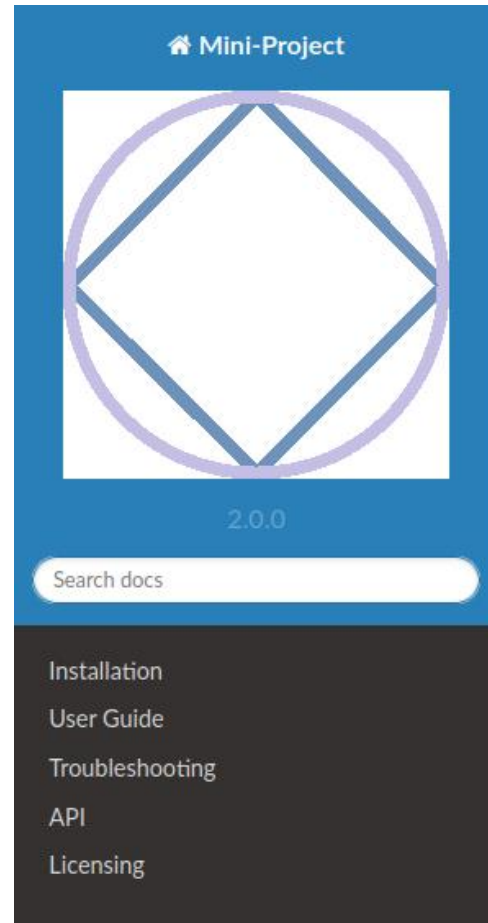
Continuous integration

- Used GitHub Actions to run our CI workflow
- Allow us to check that our package is installed correctly and passes tests on a fresh environment at any update of the git code.
- Automate documentation and release under some condition.



Documentation Sphinx

- ReStructuredText
 - README
 - tab
 - DocString
- confy.py
 - Basic information
 - HTML
- templates



» Welcome to Mini-Project's documentation!

[View page source](#)

Welcome to Mini-Project's documentation!

Continuous Integration passing
coverage 95%
tag v2.0.0
docs latest
pypi project

Mini-project

This is an extensible and fully reproducible system to analyze multiple datasets, with various Machine Learning techniques.

For more guidelines and informations on the project please refer to our documentation.

Authors

Cédric Mariéthoz and Adrien Chassignet

Documentation

Documentation HTML

Index

Installation

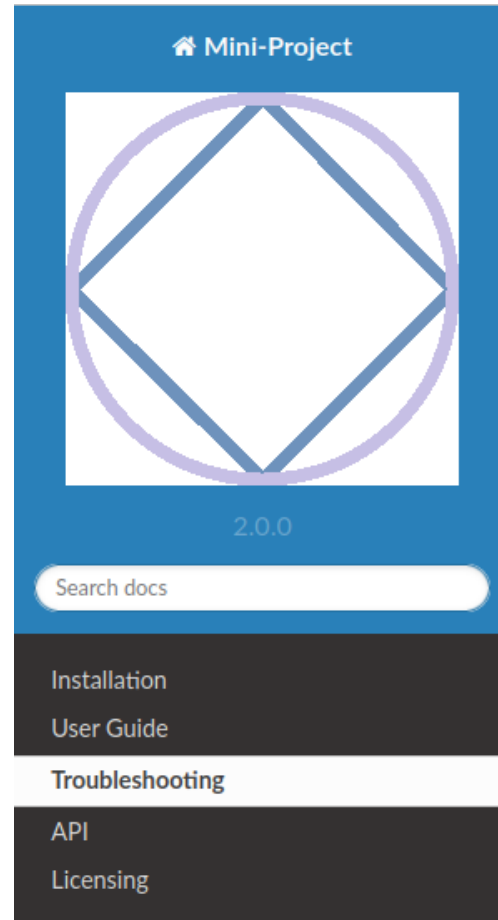
User Guide

Troubleshooting

API

License

Links



» Troubleshooting

[View page source](#)

Troubleshooting

To run the unit test we use `pytest`:

```
# use your package manager to install the package "pytest"
# here, I exemplify with "pip":
$ pip install pytest
$ pytest
....
===== test session starts =====
...
===== 40 passed, 2 skipped, 13 warnings in 8.66s =====
```

If you want to know the coverage of test, you can use this command:

```
$ pytest --cov=.

===== test session starts =====
...
----- coverage: platform linux, python 3.8.8-final-0 -----
```

TestPyPI

- requierment.txt
- setup.py
 - version
 - author
- MANIFEST.in
 - Datas

Release history [Release notifications](#) | [RSS feed](#)

| Version | Status | Time |
|---------|--------------|-------------------|
| 2.0.0 | THIS VERSION | 7 minutes ago |
| 2.0.0a1 | PRE-RELEASE | about 3 hours ago |
| 2.0.0a0 | PRE-RELEASE | Apr 6, 2022 |
| 1.0.1 | YANKED | Apr 5, 2022 |
| 1.0.1a1 | PRE-RELEASE | Apr 5, 2022 |
| 1.0.0 | YANKED | Apr 4, 2022 |

License

BSD-3

BSD

