SciKit Data Documentation

Release 0.1.2

Ivan Ogasawara

Jul 25, 2017

Contents

Contents:

CHAPTER 1

SciKit Data

Conda package current release info

About SciKit Data

The propose of this library is to allow the data analysis process more easy and automatic.

This library is based on some important libraries as:

- pandas;
- jupyter;
- matplotlib;
- scikit-learn;
- Free software: MIT license
- Documentation: https://skdata.readthedocs.io.

Features

Books used as reference to guide this project:

- https://www.packtpub.com/big-data-and-business-intelligence/clean-data
- https://www.packtpub.com/big-data-and-business-intelligence/python-data-analysis
- https://www.packtpub.com/big-data-and-business-intelligence/mastering-machine-learning-scikit-learn

Some other materials used as reference:

- https://github.com/rsouza/MMD/blob/master/notebooks/3.1_Kaggle_Titanic.ipynb
- https://github.com/agconti/kaggle-titanic/blob/master/Titanic.ipynb
- https://github.com/donnemartin/data-science-ipython-notebooks/blob/master/kaggle/titanic.ipynb

This project contemplates the follow features:

- Data conversions:
  - soon ...
- Data collection:
  - soon ...
- Data cleaning:
  - ...
- Data storage:
  - soon ...
- Data integration:
  - soon ...
- Data manipulation:
  - ...
- Outliers removal:
  - ...

Installing scikit-data

Using conda

Installing *scikit-data* from the *conda-forge* channel can be achieved by adding *conda-forge* to your channels with:

```
$ conda config --add channels conda-forge
```

Once the *conda-forge* channel has been enabled, *scikit-data* can be installed with:

```
$ conda install scikit-data
```

It is possible to list all of the versions of *scikit-data* available on your platform with:

```
$ conda search scikit-data --channel conda-forge
```

Using pip

To install scikit-data, run this command in your terminal:

```
$ pip install skdata
```

If you don't have pip installed, this Python installation guide can guide you through the process.

CHAPTER 2

Installation

Using conda

Installing *scikit-data* from the *conda-forge* channel can be achieved by adding *conda-forge* to your channels with:

```
$ conda config --add channels conda-forge
```

Once the *conda-forge* channel has been enabled, *scikit-data* can be installed with:

```
$ conda install scikit-data
```

It is possible to list all of the versions of *scikit-data* available on your platform with:

```
$ conda search scikit-data --channel conda-forge
```

Using pip

To install scikit-data, run this command in your terminal:

```
$ pip install skdata
```

If you don't have pip installed, this Python installation guide can guide you through the process.

From sources

The sources for scikit-data can be downloaded from the Github repo.

You can either clone the public repository:

```
$ git clone git://github.com/OpenDataScienceLab/skdata
```

Or download the tarball:

```
$ curl  -OL https://github.com/OpenDataScienceLab/skdata/tarball/master
```

Once you have a copy of the source, you can install it with:

```
$ python setup.py install
```

CHAPTER 3

Usage

To use SciKit Data Analisys in a project:

```
import skdata
```

CHAPTER 4

Contributing

Contributions are welcome, and they are greatly appreciated! Every little bit helps, and credit will always be given.

You can contribute in many ways:

Types of Contributions

Report Bugs

Report bugs at https://github.com/OpenDataScienceLab/skdata/issues.

If you are reporting a bug, please include:

- Your operating system name and version.
- Any details about your local setup that might be helpful in troubleshooting.
- Detailed steps to reproduce the bug.

Fix Bugs

Look through the GitHub issues for bugs. Anything tagged with "bug" and "help wanted" is open to whoever wants to implement it.

Implement Features

Look through the GitHub issues for features. Anything tagged with "enhancement" and "help wanted" is open to whoever wants to implement it.

Write Documentation

Jupyter Python Data Analisys could always use more documentation, whether as part of the official Jupyter Python Data Analisys docs, in docstrings, or even on the web in blog posts, articles, and such.

Submit Feedback

The best way to send feedback is to file an issue at https://github.com/xmnlab/skdata/issues.

If you are proposing a feature:

- Explain in detail how it would work.
- Keep the scope as narrow as possible, to make it easier to implement.
- Remember that this is a volunteer-driven project, and that contributions are welcome :)

Get Started!

Ready to contribute? Here's how to set up *skdata* for local development.

1. Fork the *skdata* repo on GitHub.
2. Clone your fork locally:

```
$ git clone git@github.com:your_name_here/skdata.git
```

3. Install your local copy into a virtualenv. Assuming you have virtualenvwrapper installed, this is how you set up your fork for local development:

```
$ mkvirtualenv skdata
$ cd skdata/
$ python setup.py develop
```

4. Create a branch for local development:

```
$ git checkout -b name-of-your-bugfix-or-feature
```

   Now you can make your changes locally.

5. When you're done making changes, check that your changes pass flake8 and the tests, including testing other Python versions with tox:

```
$ flake8 skdata tests
$ python setup.py test or py.test
$ tox
```

   To get flake8 and tox, just pip install them into your virtualenv.

6. Commit your changes and push your branch to GitHub:

```
$ git add .
$ git commit -m "Your detailed description of your changes."
$ git push origin name-of-your-bugfix-or-feature
```

7. Submit a pull request through the GitHub website.

Pull Request Guidelines

Before you submit a pull request, check that it meets these guidelines:

1. The pull request should include tests.
2. If the pull request adds functionality, the docs should be updated. Put your new functionality into a function with a docstring, and add the feature to the list in README.rst.
3. The pull request should work for Python 3.4 and 3.5. Check https://travis-ci.org/xmnlab/skdata/pull_requests and make sure that the tests pass for all supported Python versions.

Tips

To run a subset of tests:

```
$ python -m unittest tests.test_skdata
```

CHAPTER 5

Indices and tables

- genindex
- modindex
- search