



# Training with open-source

Tom Dupré la Tour

PyData Paris - 14-15 June 2016



# Who am I?

- My experience with **scikit-learn**:
  - 5 months – full-time
  - 1 year – 1/6 time during my PhD – “Open Software Initiative” (Paris Saclay Center for Data science)

# Who am I?

- My experience with **scikit-learn**:
  - 5 months – full-time
  - 1 year – 1/6 time during my PhD – “Open Software Initiative” (Paris Saclay Center for Data science)
- Working on:
  - Algorithm implementation (SAG, CD-NMF, MU-NMF)
  - Maintenance: debug, documentation, testing, refactoring, ...
  - Code reviews, help for newcomers

# Why did I need a training?

- Education:
  - 5 years: Preparatory school + Engineering school
  - 1.5 year: Master degree in Switzerland

# Why did I need a training?

- Education:
  - 5 years: Preparatory school + Engineering school
  - 1.5 year: Master degree in Switzerland
- What I knew:
  - Mathematics, Physics, Mechanics, Electronics, Algorithmics, Economics, ...
  - Java, Matlab
  - Dropbox

# Why did I need a training?

- Education:
  - 5 years: Preparatory school + Engineering school
  - 1.5 year: Master degree in Switzerland
- What I knew:
  - Mathematics, Physics, Mechanics, Electronics, Algorithmics, Economics, ...
  - Java, Matlab
  - Dropbox
- What I (almost) didn't know:
  - Git, GitHub, Linux, Python, ...
  - Unit testing, Continuation Integration, Coding standards, ...



# How does open-source help?

- Mentoring:
  - Code review
  - Programming sessions (pair programming, coding sprint)

# How does open-source help?

- Mentoring:
  - Code review
  - Programming sessions (pair programming, coding sprint)
- High quality standards:
  - Read high-quality code (PEP8, API design, GitHub discussions)
  - Read high-quality documentation



# How does open-source help?


- Mentoring:
  - Code review
  - Programming sessions (pair programming, coding sprint)
- High quality standards:
  - Read high-quality code (PEP8, API design, GitHub discussions)
  - Read high-quality documentation
- A whole new world:
  - GitHub, Blogs, Twitter, ...
  - Super dynamic ecosystem

  **TomDLT** commented on an outdated diff 12 days ago

 Show 1 comment

  **TomDLT** commented on an outdated diff 12 days ago



 Show 1 comment

  **TomDLT** commented on an outdated diff 12 days ago



 Show 1 comment

  **TomDLT** commented on an outdated diff 12 days ago













 Show 1 comment




  **TomDLT** commented on an outdated diff 12 days ago

 Show 1 comment

  **TomDLT** and 1 other commented on an outdated diff 12 days ago

 Show 2 comments

-   **Jnothman** commented on an outdated diff 6 days ago  Show 1 comment
-   **Jnothman** commented on an outdated diff 6 days ago  Show 1 comment
-   **Jnothman** commented on an outdated diff 6 days ago  Show 1 comment
-   **Jnothman** commented on an outdated diff 6 days ago  Show 1 comment


  **TomDLT** and 1 other commented on an outdated diff 4 days ago  Show 2 comments

  **agramfort** commented on an outdated diff 4 days ago  Show 1 comment

  **agramfort** commented on an outdated diff 4 days ago  Show 1 comment

  **agramfort** commented on an outdated diff 4 days ago  Show 1 comment

  **agramfort** commented on an outdated diff 4 days ago  Show 1 comment

  **agramfort** commented on an outdated diff 4 days ago  Show 1 comment

  **agramfort** commented on an outdated diff 4 days ago  Show 1 comment

  **agramfort** commented on an outdated diff 4 days ago  Show 1 comment

  **agramfort** commented on an outdated diff 4 days ago  Show 1 comment



**lesteve** commented on an outdated diff a day ago



Show 1 comment



**lesteve** commented on an outdated diff a day ago



Show 1 comment



**lesteve** commented on an outdated diff a day ago



Show 1 comment



**jnothman** commented on an outdated diff a day ago




Show 1 comment






**jnothman** commented on an outdated diff a day ago






Show 1 comment

  **lesteve** commented on an outdated diff a day ago  Show 1 comment



---

  **lesteve** commented on an outdated diff a day ago  Show 1 comment

---

  **lesteve** commented on an outdated diff a day ago  Show 1 comment


---

  **jnothman** commented on an outdated diff a day ago  Show 1 comment

---

  **jnothman** commented on an outdated diff a day ago  Show 1 comment

---

  **jnothman** commented on an outdated diff an hour ago  Show 2 comments

---

# Is this training efficient?

- Fast learning curve:
  - You don't learn code in books: code, and get feedbacks
  - High-quality feedbacks
  - From different people
  - Fast sedimentation (realized with newcomers)

# Is this training efficient?

- Fast learning curve:
  - You don't learn code in books: code, and get feedbacks
  - High-quality feedbacks
  - From different people
  - Fast sedimentation (realized with newcomers)
- Most of the time, in parallel of your job/education:
  - 700+ contributors
  - Many students, researchers, developers



# So what did I learn?

- The Python way:
  - Clarity, importance of documentation, of tests, of comments
  - The code is made to be re-used/re-written by many

# So what did I learn?

- The Python way:
  - Clarity, importance of documentation, of tests, of comments
  - The code is made to be re-used/re-written by many
- Coding for my PhD:
  - Mathematics/Neuroscience? Most of my work relies on code
  - Modularity, API design, efficiency

# So what did I learn?

- The Python way:
  - Clarity, importance of documentation, of tests, of comments
  - The code is made to be re-used/re-written by many
- Coding for my PhD:
  - Mathematics/Neuroscience? Most of my work relies on code
  - Modularity, API design, efficiency
- Machine Learning:
  - Excellent overview of estimators, methodology, pipelines, ...

# Is that all?

- Collaborative experience:
  - With passionate people
  - Open-source values

# Is that all?

- Collaborative experience:
  - With passionate people
  - Open-source values
- Used by many:
  - Rewarding work
  - Good visibility (for both academic and industry)

Thank you for your attention