

# Lessons learned from *scikit-learn* (and how they helped me in research & life)

Vlad Niculae  
Cornell University

<https://vene.ro>



# Who's this guy?

Sterling is much cheaper than gold.

vs.

Her voice makes this song shine brighter than gold.

<https://vene.ro/figurative-comparisons/>

# Who's this guy?

America has been through big changes before—wars and depression, the influx of new immigrants, workers fighting for a fair deal [...]

**The New York Times**

<http://snap.stanford.edu/quotus/>



# Who's this guy?

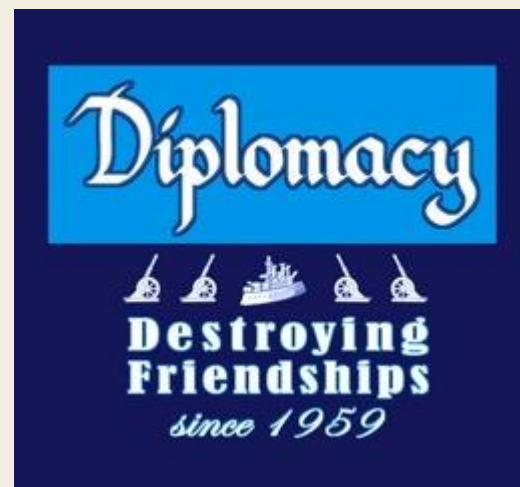
America has been through big changes before—wars and depression, the influx of new immigrants, ~~workers fighting for a fair deal [...]~~

<http://snap.stanford.edu/quotus/>



# Who's this guy?

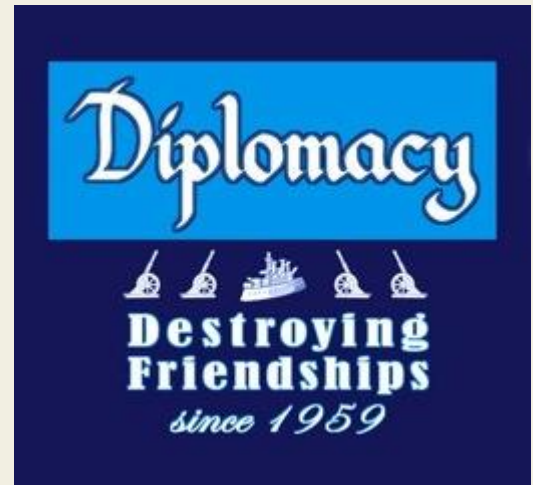
Would it be ok with you if I took  
Denmark? I think I'm going to need it  
if I am going to hold France back.



<https://vene.ro/betrayal>

# Who's this guy?

Would it be a good idea if I took  
Denmark's place? I'm going to need it  
if I am going to come back.



<https://vene.ro/betrayal>

# Who's this guy?



CMV: Martin Shkreli is a victim of the uninformed and extremely biased media.

<https://chenhaot.com/pages/changemyview.html>

- Why contribute to open source?
- Why is scikit-learn successful?
- What are good practices for research code?



- Why contribute to open source?  
Huge learning potential.
- Why is scikit-learn successful?
- What are good practices for research code?

- Why contribute to open source?  
Huge learning potential.
- Why is scikit-learn successful?  
Good, consistent design and principles.
- What are good practices for research code?

- Why contribute to open source?  
Huge learning potential.
- Why is scikit-learn successful?  
Good, consistent design and principles.
- What are good practices for research code?  
Lessons from libraries transfer well.

# Why contribute?

(among many reasons)

- Great feeling when people use your stuff. **BSD!**
- Have code reviewed & review great code by others.

Not just coding tricks!

Glimpse into what others use and why.

- Start seeing your code through an API lens.
- Join a [community](#), blog, get retweets.

The secret sauce (user's angle)

# The secret sauce (user's angle)

- Not a *framework*:  
minimal glue, zero magic



# The secret sauce (user's angle)

- Not a *framework*:  
minimal glue, zero magic



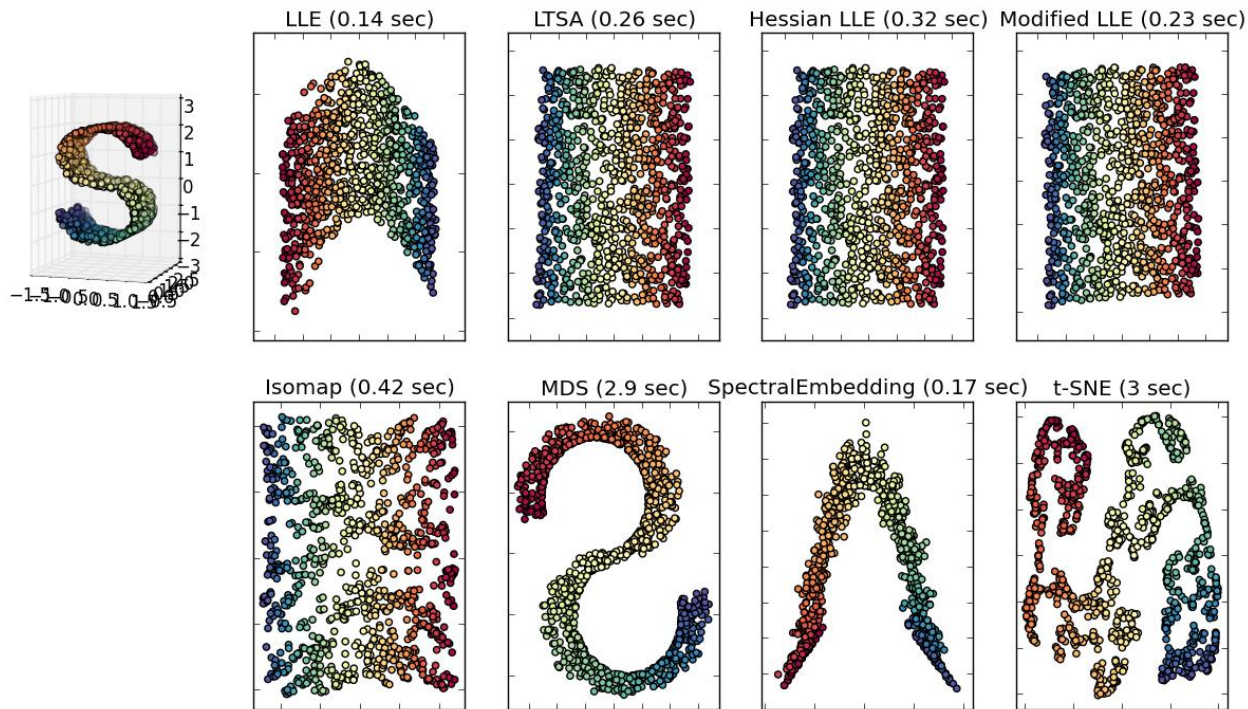
# The secret sauce (user's angle)

- Not a *framework*:  
minimal glue, zero magic
- Documentation, motivating examples.



# The secret sauce (user's angle)

Manifold Learning with 1000 points, 10 neighbors



# The secret sauce (user's angle)

- Not a *framework*:  
minimal glue, zero magic
- Documentation, motivating examples.
- Honest choice of what algorithms to include.

# The secret sauce (user's angle)

- Not a *framework*:  
minimal glue, zero magic
- Documentation, motivating examples.
- Honest choice of what algorithms to include.

Learning the parts of objects by **non-negative matrix factorization**

[DD Lee](#), [HS Seung](#) - Nature, 1999 - [nature.com](#)

Abstract Is perception of the whole based on perception of its parts? There is psychological 1 and physiological 2, 3 evidence for parts-based representations in the brain, and certain computational theories of object recognition rely on such representations 4, 5. But little is ...

Cited by 6311   Related articles   All 20 versions   Cite   Save

# The secret sauce (user's angle)

- Not a *framework*:  
minimal glue, zero magic
- Documentation, motivating examples.
- Honest choice of what algorithms to include.

Learning the parts of objects by **non-negative matrix factorization**

[DD Lee](#), [HS Seung](#) - Nature, 1999 - [nature.com](#)

Abstract Is perce  
1 and physiologi  
computational th

Cited by 6311 f

Super Deep Sparse Extreme Machines **[2016 FRESH MEGAHIT!!]**

John Doe, arXiv preprint

Abstract Is perception of the whole based on perception of its parts? I here is psychological  
1 and physiological 2, 3 evidence for parts-based representations in the brain, and certain  
computational theories of object recognition rely on such representations 4, 5. But little is ...

Cited by **-3** Related articles All 20 versions Cite Save

# The secret sauce (user's angle)

- Not a *framework*:  
minimal glue, zero magic
- Documentation, motivating examples.
- Honest choice of what algorithms to include.

Learning the parts of objects by non-negative matrix factorization

DD Lee, HS Seung - Nature, 1999 - nature.com

Abstract Is perce  
1 and physiologi  
computational th

Cited by 6311 f

Super Deep Sparse Extreme Machines [2016 FRESH MEGAHIT!!]

John **Library-Author**, arXiv preprint

Abstract Is perception of the whole based on perception of its parts? There is psychological  
1 and physiological 2, 3 evidence for parts-based representations in the brain, and certain  
computational theories of object recognition rely on such representations 4, 5. But little is ...

Cited by -3 Related articles All 20 versions Cite Save

# The secret sauce (user's angle)

- Not a *framework*:  
minimal glue, zero magic
- Documentation, motivating examples.
- Honest choice of what algorithms to include.
- Explore = develop = deploy

# The secret sauce (dev angle)

- Not a *framework*:  
Easier to debug!  
Sometimes frustratingly limiting.
- Documentation,  
motivating examples:
  - (a) force devs to actually think more
  - (b) good filter against bloat, useless contribs
- Test coverage.



## 4 The tragedy of the commons

*Individuals, acting independently and rationally according to each one's self-interest, behave contrary to the whole group's long-term best interests by depleting some common resource.*

Wikipedia



**Make it work, make it right, make it boring**

Core projects (boring) taken for granted

⇒ Hard to fund, less excitement

**They need citation, in papers & on corporate web pages**



# In your own work:

*but don't look at mine :)*

- Your (research) code is a library, and your future self 3 weeks from now is a potential user.

Make collaborators review your research code!

- Code without **unit tests** is broken.  
Code with unit tests at least works sometimes.

This is at odds with *get it done...*

- Everything is useless until shown useful.

Don't get caught up  
on the latest greatest hyped method.