Nottinghamshire Healthcare
NHS Foundation Trust

R and Python – a happy union with reticulate
Part B

Andreas Soteriades, Data Scientist (CDU Data Science Team, Nottinghamshire Healthcare NHS Foundation Trust)

NHS-R Community, 20/01/2021

Making a Difference
Trust Honesty Respect Compassion Teamwork

Scikit-learn (

)

- Scikit-learn is probably the most popular ML library in Python.
- It offers a wealth of functions and classes for (un)supervised learning, feature engineering & selection, hyperparameter tuning etc.
- It does not have Deep Learning models, but integration with e.g. Keras is easily done with a wrapper.
- It has several advantages on its own, but also relative to key ML packages in R like tidymodels and mlr3. Let's discover them!

Scikit-learn: Machine Learning in Python, Pedregosa *et al.*, JMLR 12, pp. 2825-2830, 2011.

Scikit-learn & its R counterparts

Nottinghamshire Healthcare NHS Foundation Trust

	R		Python
	tidymodels	**mlr3**	**Scikit-learn**
ML models	All three libraries offer an interface for integrating and standardizing the use of different models. The problem with R though is that it borrows models from different packages, the quality of which depends on the authors' skills and individual efforts (e.g. willingness to make models faster, add more features or actively maintain package). In Scikit-learn, models in it are built *for* it.		
Speed	I am not aware of a benchmarking exercise. But mlr3 could be faster and more efficient, because it uses data.table and R6 objects. Speed also strongly depends on the individual packages that tidymodels and mlr3 borrow the ML models from.		Designed to be fast. Interoperates with NumPy and SciPy for fast scientific computing. Many core algorithms built in Cython.
User-friendliness	tidyverse-style use of "%>%". Loads of resources. More appropriate for newbies.	Resources a little scattered and some outdated. More "hardcore" ML, but easy to catch up if familiar with ML, Python and "classes".	Solid, well-organized and consistent User Guide that also covers model theory and has numerous examples.
Text classification	Possible, with a few models available for Bag of Words (BoW) learning-although I do not know how fast.	Currently at early stage and slow.	Large collection of mind-blowingly fast models for BoW.

The verdict

Nottinghamshire Healthcare
NHS Foundation Trust

tidymodels will help you to:

- Familiarize yourselves with ML, pipelines, pre-processing, benchmarking etc.
- Practice on smaller tasks.

The verdict

Nottinghamshire Healthcare
NHS Foundation Trust

tidymodels will help you to:

- Familiarize yourselves with ML, pipelines, pre-processing, benchmarking etc.
- Practice on smaller tasks.

mlr3 will help you to:

- Understand (or at least get used to) classes & pipelines that closely resemble the structure in Scikit-learn.
- Run bigger and more complicated tasks efficiently.

The verdict

Nottinghamshire Healthcare
NHS Foundation Trust

tidymodels will help you to:

- Familiarize yourselves with ML, pipelines, pre-processing, benchmarking etc.
- Practice on smaller tasks.

mlr3 will help you to:

- Understand (or at least get used to) classes & pipelines that closely resemble the structure in Scikit-learn.
- Run bigger and more complicated tasks efficiently.

Scikit-learn will help you to:

- Delve deeper into the world of ML and get things done fast.
- Enter the world of Python, where ML is huge (TensorFlow, Keras, spaCy, VADER, etc.).

The verdict

Nottinghamshire Healthcare NHS Foundation Trust

tidymodels will help you to:

- Familiarize yourselves with ML, pipelines, pre-processing, benchmarking etc.
- Practice on real data.

mlr3 will help you to:

- Understand (or at least get used to) classes & pipelines that closely resemble the structure in Scikit-learn.
- Run bigger and more complicated tasks efficiently.

Scikit-learn will help you to:

- Delve deeper into the world of ML and get things done fast.
- Enter the world of Python, where ML is huge (TensorFlow, Keras, spaCy, VADER, etc.).

Python's ML wins!

"But I'm 100% an R user… What can I do?"

The verdict

Nottinghamshire Healthcare NHS Foundation Trust

tidymodels will help you to:

- Familiarize yourselves with ML, pipelines, pre-processing, benchmarking etc.
- Practice on [illegible]

mlr3 will help you to:

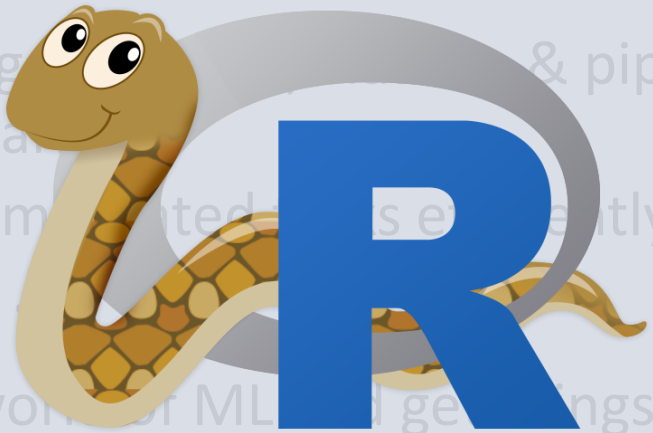
- Understand (or at least g[illegible] & pipelines that closely resemble the structure in Scikit-lea[illegible]
- Run bigger and more com[illegible]ted[illegible]s e[illegible]tly.

Scikit-learn will help you [illegible]

- Delve deeper into the wo[illegible] of ML[illegible] ge[illegible]ngs done fast.
- Enter the world of Python, where ML is huge (TensorFlow, Keras, spaCy, VADER, etc.).

Python's ML wins!

"But I'm 100% an R user… What can I do?"



★**Reticulate!**★