

Data Analysis methods that are changing (CEDAR) science

Karthik Venkataramani

ASTRA

06/23/2020

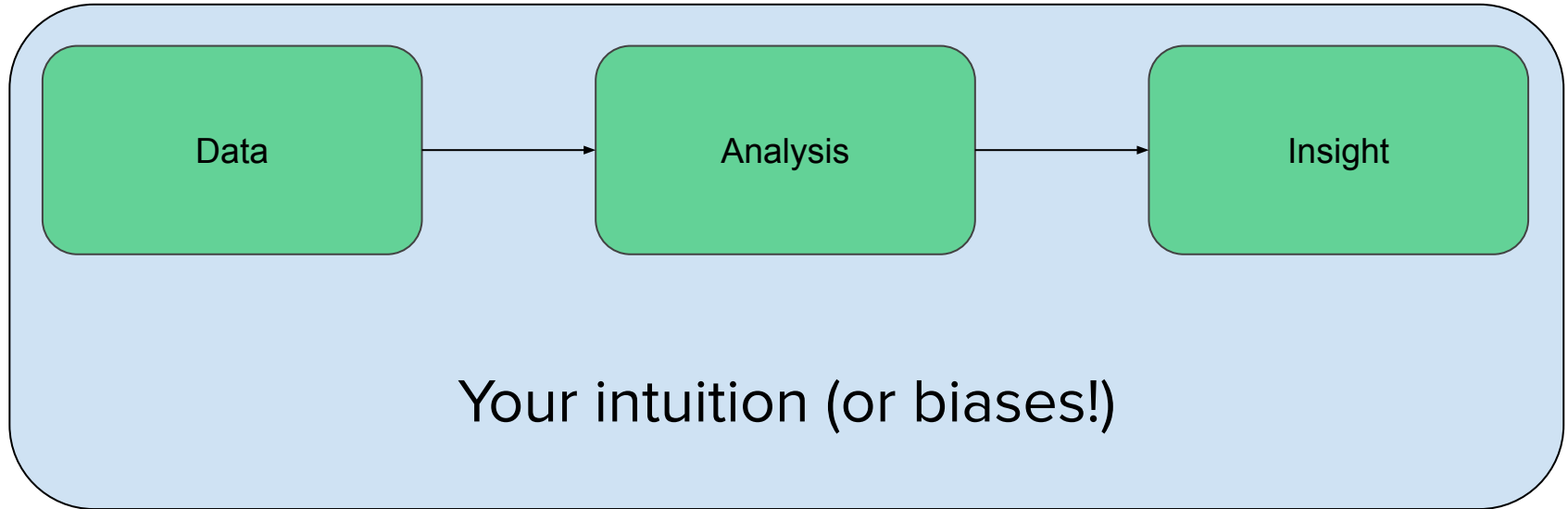


kvenkman

Approaching the problem



Approaching the problem



What's new?

Choice of tools is often crucial

The usual suspects -

- Python (NumPy, Scikit-learn, Pandas ..) [[Sorry IDL!](#)]
- Visualization tools (Matplotlib, Bokeh, Seaborn)
- Jupyter Notebooks
- Git / GitHub
- Twitter?!

Check out the '[Snakes on a spaceship](#)' workshop!

Demo time!

Git is intimidating, but ..

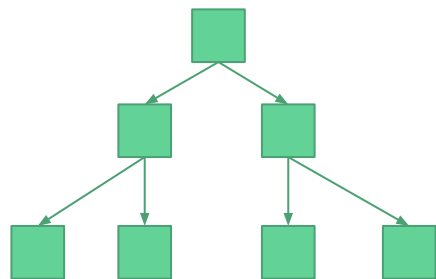
There are plenty of resources online to learn, and the best way to learn is by asking questions!

My personal favorite resource: <https://dangitgit.com/>

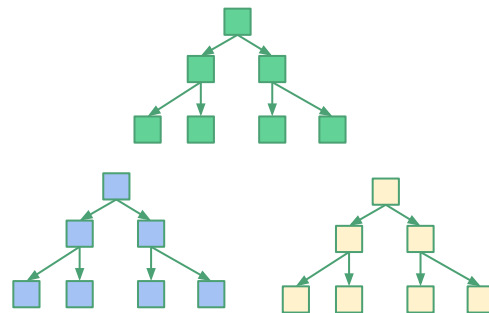
Learning from the Machines

Ensemble Learning

“Random decision forests correct for decision trees' habit of overfitting to their training set” [1]



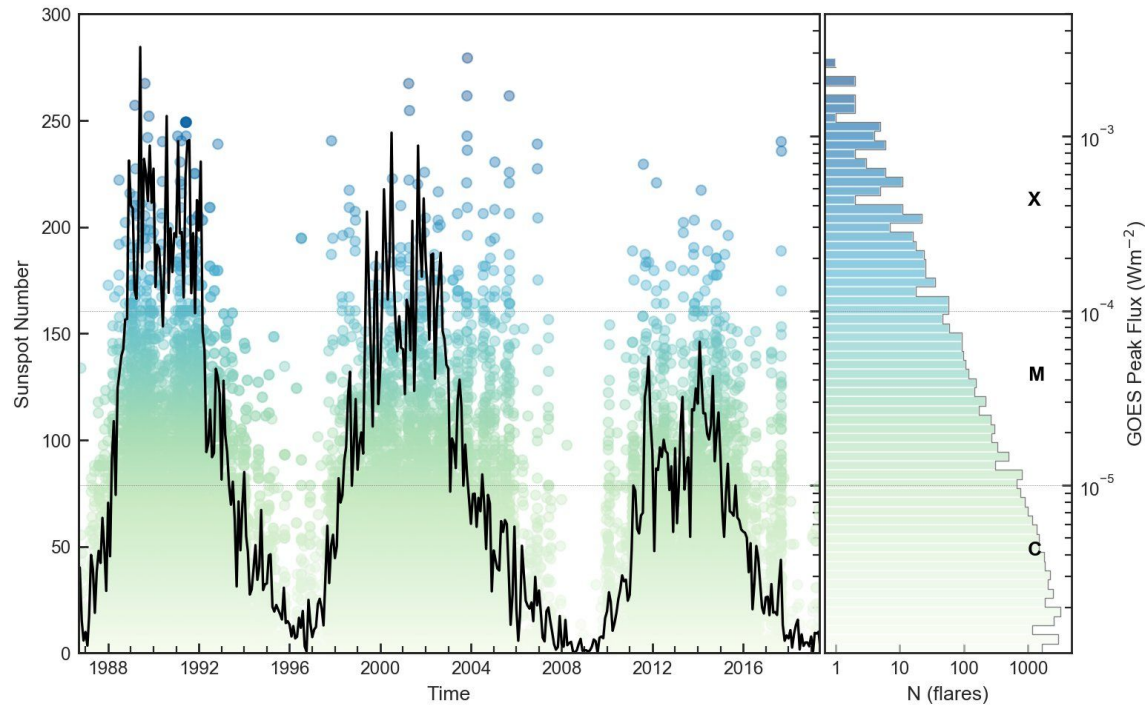
A Decision Tree
(You)



A Random Forest
(You + Collaborators!)

[1] “*The elements of statistical learning.*” Friedman, Jerome, Trevor Hastie, and Robert Tibshirani.

Twitter?



Twitter!

