Data Science at LHCb
The Large Hadron Collider
This is the right time:

Dark Matter, stuff vs anti-stuff
New Physics 101
How many are there?
Who ya gonna call?
Cherenkov - faster than the speed of light
How many are there?
What else does the LHC produce?
Simulation

By Joel Holdsworth (Joelholdsworth) CC-BY-SA-3.0
Simulate all the things
Simulate all the things #hard
What else does the LHC produce?
Discard events as early as possible
Real Time Stream Processing
“Real” Time Stream Processing
The Topological Trigger

Gligorov, Williams, arXiv:1210.6861
“I’ve mostly figured out what the BDT does. In easy regions it does easy to understand things and in hard regions it does hard to understand things.”
The Topological Trigger

Old
Fast
Robust
General purpose

Gligorov, Williams, arXiv:1210.6861
The Final Selection
The Final Selection

This is where kaggle starts
Power Users

Performance!!

Advanced techniques
Normal People

What is it doing?
Artisans
Artisans

Uniformity, or other special loss functions
The Final Selection
Building Models
The Known Unknowns
Publish!

CMS and LHCb collaborations, doi:10.1038/nature14474, arXiv:1411.4413
Preempting the Trigger

Learn the decision function of the whole LHCb trigger
Preempting the Trigger

No work is faster than some work.
Preempting the Trigger

Benbouzid, Busa-Fekete, Kegl, arXiv:1206.6387
Generative Models

Simulation is the largest user of CPU time!
Track Finding

Reconstruction
Extracting properties
Details matter
Tim Head

✉️ thead@cern.ch

🐦 @betatim