Prospective du LPC

Machine Learning

11 July 2018, Domaine du Marand

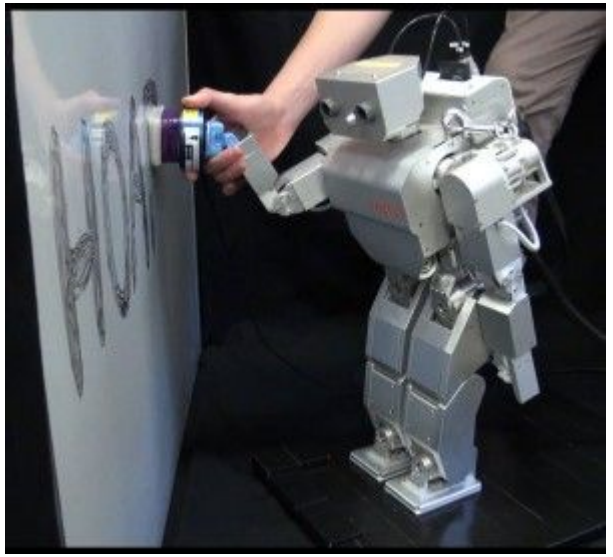
Emille E. O. Ishida
Laboratoire de Physique de Clermont - Université Clermont-Auvergne
Clermont Ferrand, France

Big Data is not more of the same ...

https://digital4africa.com/2018/02/28/whats-big-deal-data/big_data_2_neu/

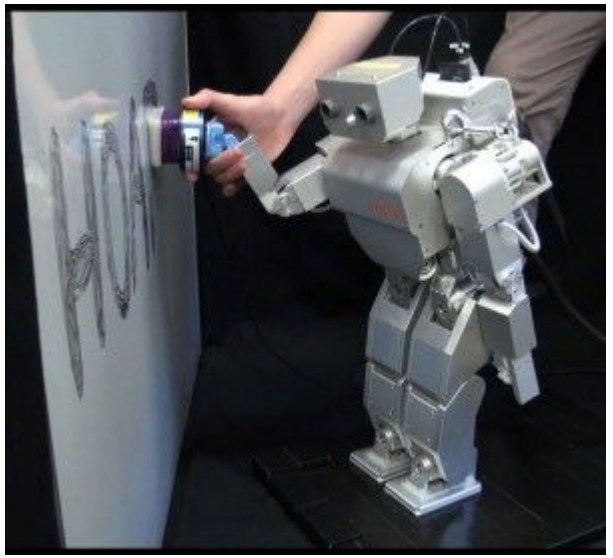
Machines can help...

… but we need to teach them!



Machines can help...

- Repetitive tasks
- Complex tasks
- New insights



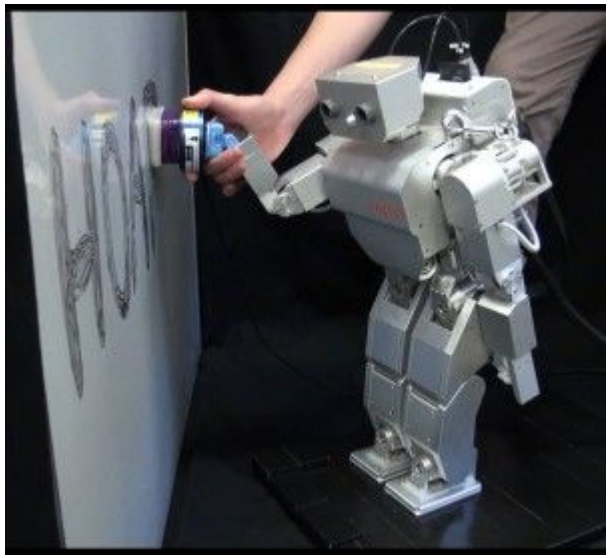
… but we need to teach them!

Machines can help...

- Repetitive tasks
- Complex tasks
- New insights

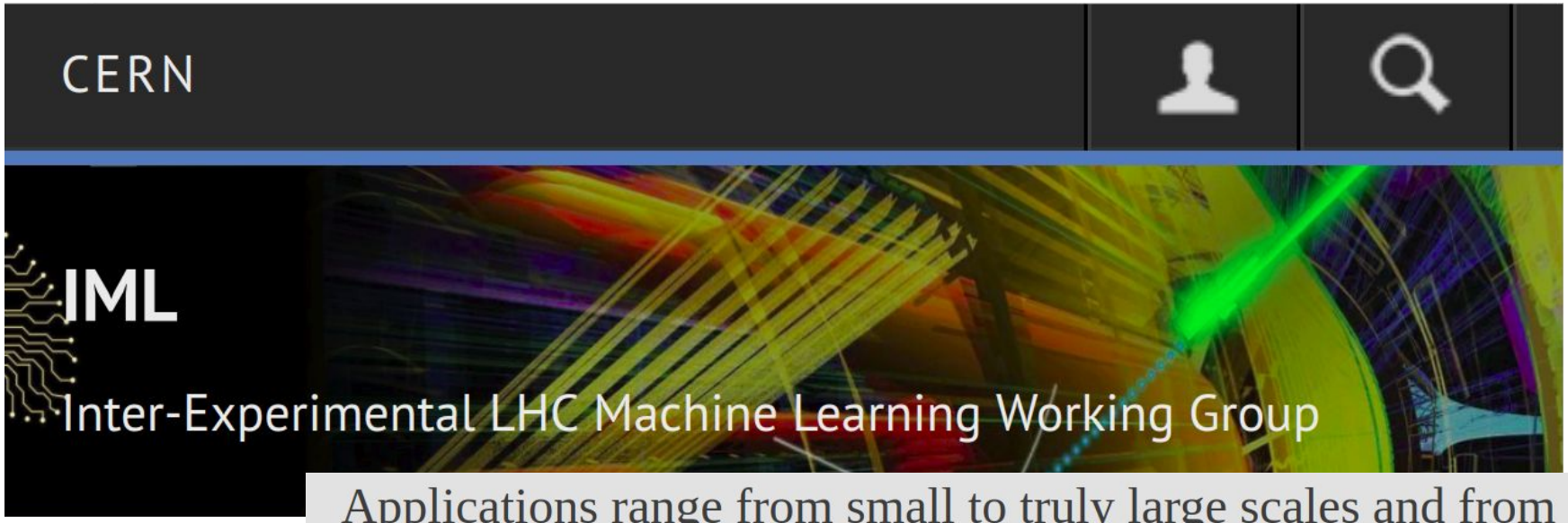
Machine Learning field needs domain knowledge to produce significant scientific results …

Even if they do not admit it!



… but we need to teach them!

Machines Learning @ CERN



Applications range from small to truly large scales and from very fast (a few μs) to modest inference (many seconds) times. The Inter-experimental Machine Learning (**IML**) Working Group provides a forum for the machine learning community at the LHC. It brings together scientists from the

https://iml.web.cern.ch/

Machine Learning in France

AstroLab Software

Bring big data and science together

Orsay, FR

Email

GitHub

AstroLab software

Developed at LAL

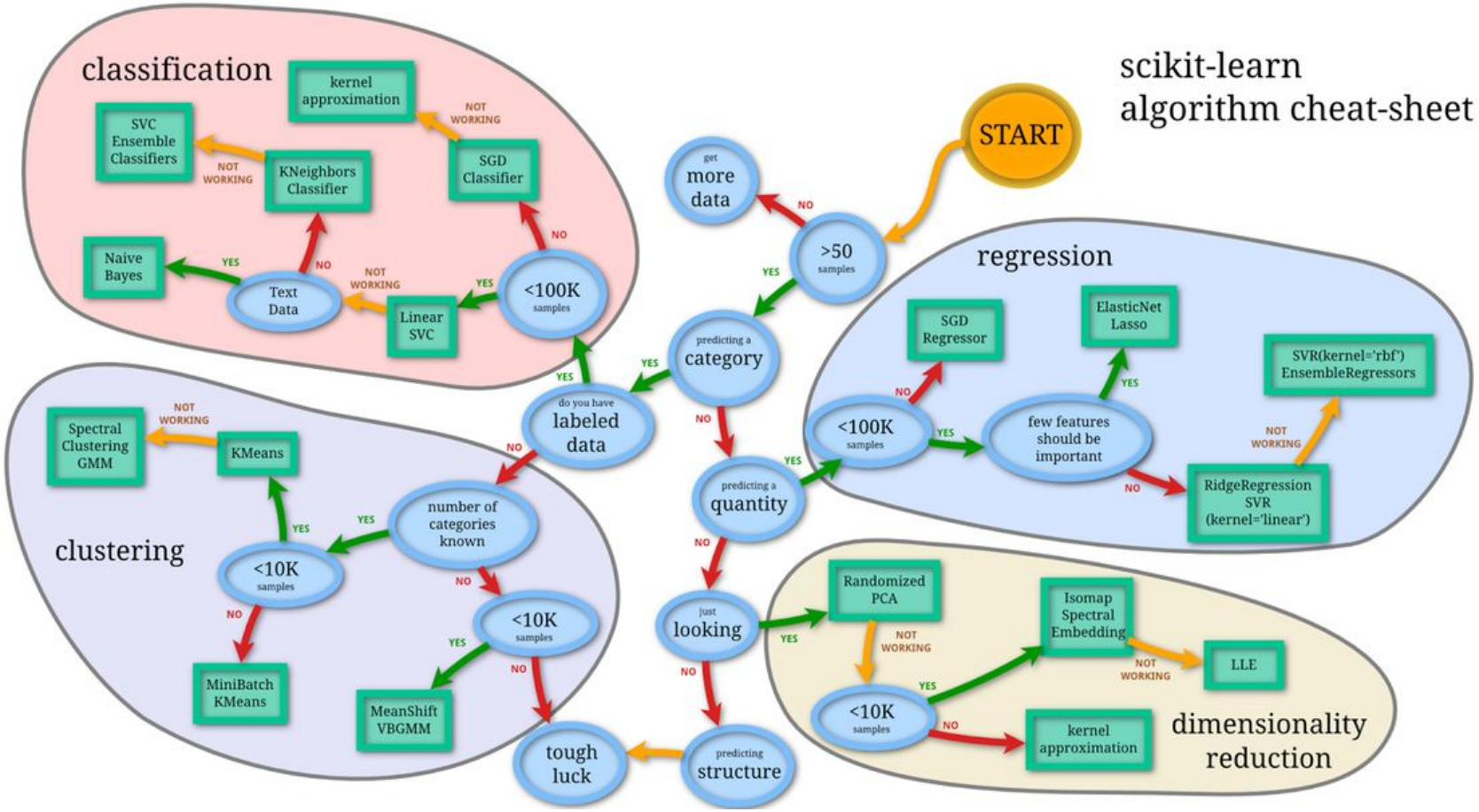
AstroLab Software is an organisation aiming at providing state-of-the-art software tools to overcome modern science challenges faced by research groups. Sharing R&D efforts between groups, improving interoperability between industry and research in open source projects, and developing new collaborative tools will allow research communities to more fully exploit the big data ecosystem tools.

https://astrolabsoftware.github.io/

Machine Learning in France

scikits learn

developed and maintained at INRIA



Machine Learning in France

Paris, le 30 mars 2018

Académiques et industriels s'unissent pour créer

l'Institut PRAIRIE[1],

lieu d'excellence dédié à l'intelligence artificielle à Paris

Machine Learning in France

A l'occasion du sommet AI for Humanity, le Président de la République Emmanuel Macron a dévoilé la stratégie française en matière d'intelligence artificielle. Il a notamment annoncé la mise en place d'un « réseau emblématique de quatre ou cinq instituts dédiés, ancrés dans des pôles universitaires et maillant le territoire ».

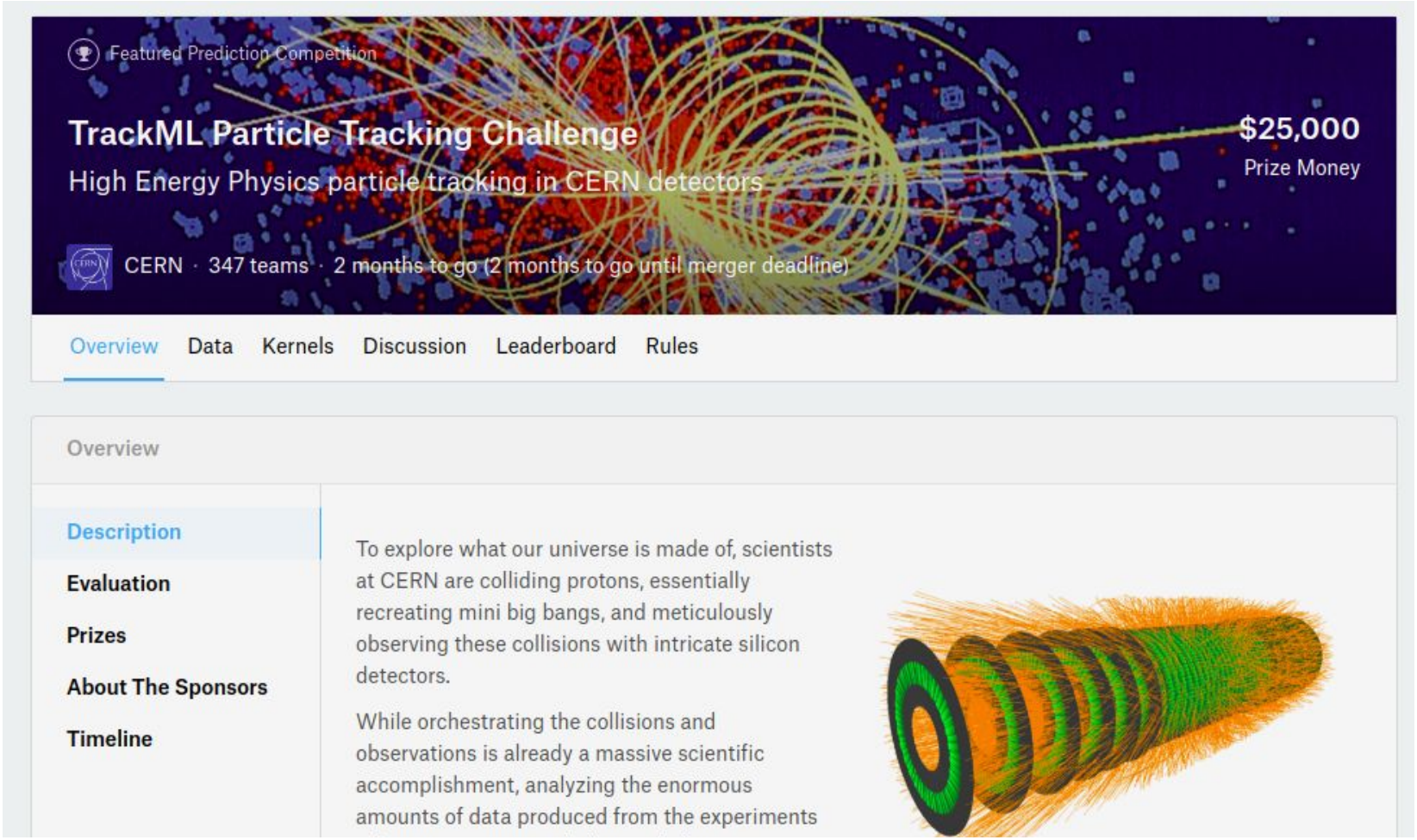
Paris, le 30 mars 2018

Académiques et industriels s'unissent pour créer

l'Institut PRAIRIE[1],

lieu d'excellence dédié à l'intelligence artificielle à Paris

TrackML - Data Challenge



https://www.kaggle.com/c/trackml-particle-identification

Machine Learning @ IN2P3

- Activité Machine Learning en cours à l'IN2P3
  - (ne compte pas l'utilisation standard de BDT, qui est toujours la technique recommandée pour classification sur une douzaine de variables)
  - generative models ATLAS calorimeter or LSST
  - anomaly detection LSST or ATLAS
  - Active Learning (LSST)
  - Active Learning for 'intelligent" simulation
  - fast tracking ATLAS LHCB
  - deblending with deep learning LSST
  - KM3net event reconstruction/identication
  - CTA event reconstruction/identification
  - Reconstruction de camera imagerie beta/gamma (medical)
  - system administration learning from log files or other information

- Pas (encore) de projet commun identifié (sauf TrackML)

4 workshops already happened

- MACHINE-LEARNING-L@IN2P3.FR : 65 participants
- MACHINE-LEARNING-CORE-TEAM-L@IN2P3.FR : Balazs Kegl, David Rousseau, Eric Aubourg, Françoise Bouvet, Emille Ishida, Emmanuel Gangler, Jérome Pansanel, Vava Gligorov Thomas Vuillaume

Discussions started because of the "Prospective du LPC":

Machine Learning @ LPC

Plan for the near future

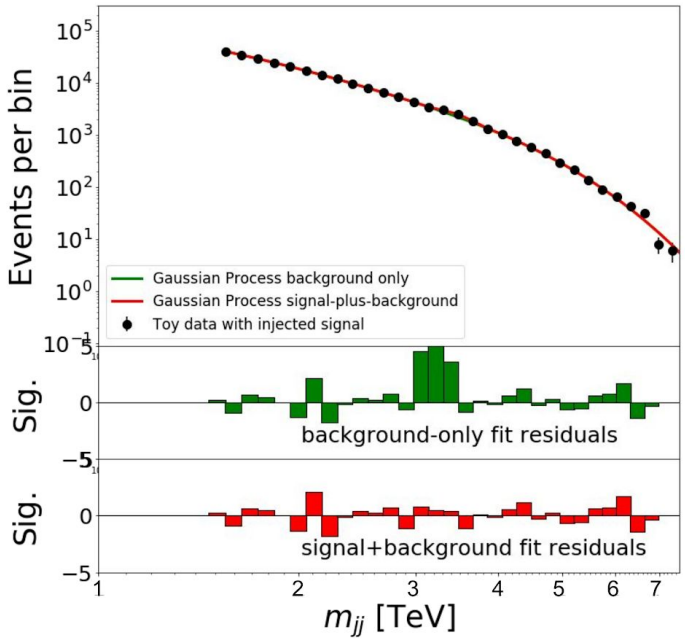
Stage 0	Stage 1	Stage 2	Stage 3
2018/1.5	2018/2	2019/1	2019/2
Know thy neighbor	Level the plainfield	Gather tools	The challenge

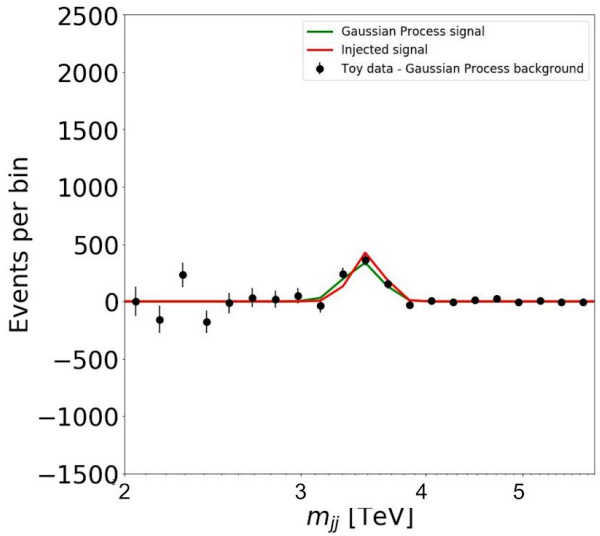
- 3 scientific meetings: April, May and June
- Between 5 to 10 participants
- 2 PhD students as speakers (Fabricio and Lennart)

Machine Learning @ LPC

Gaussian Processes for Model Independent Resonance Searches - ATLAS

2-jet mass spectrum with signal injected



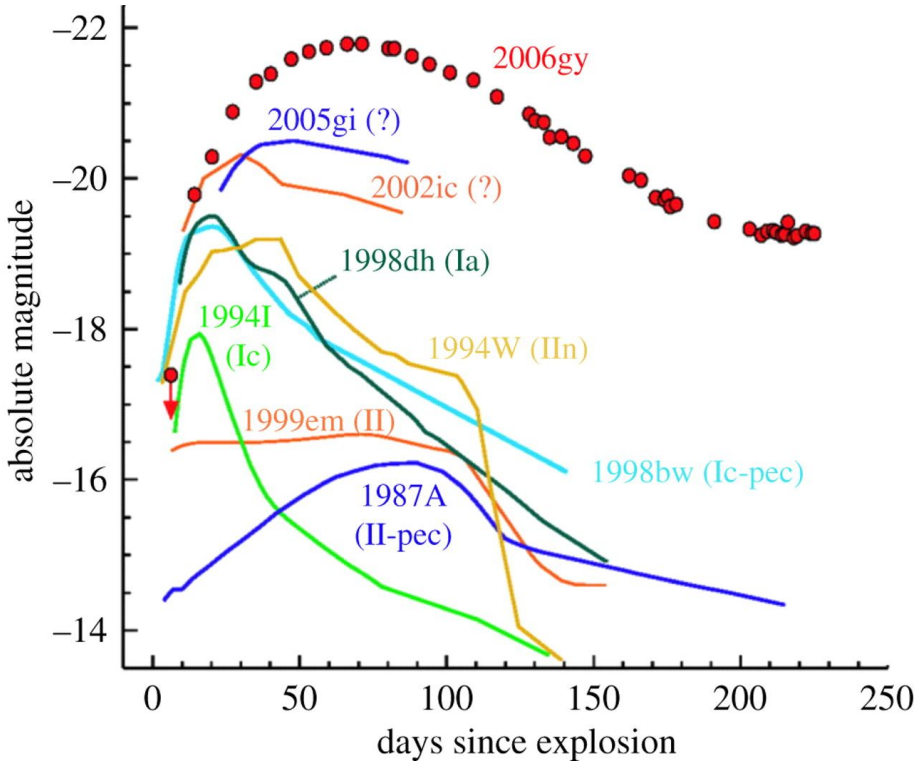


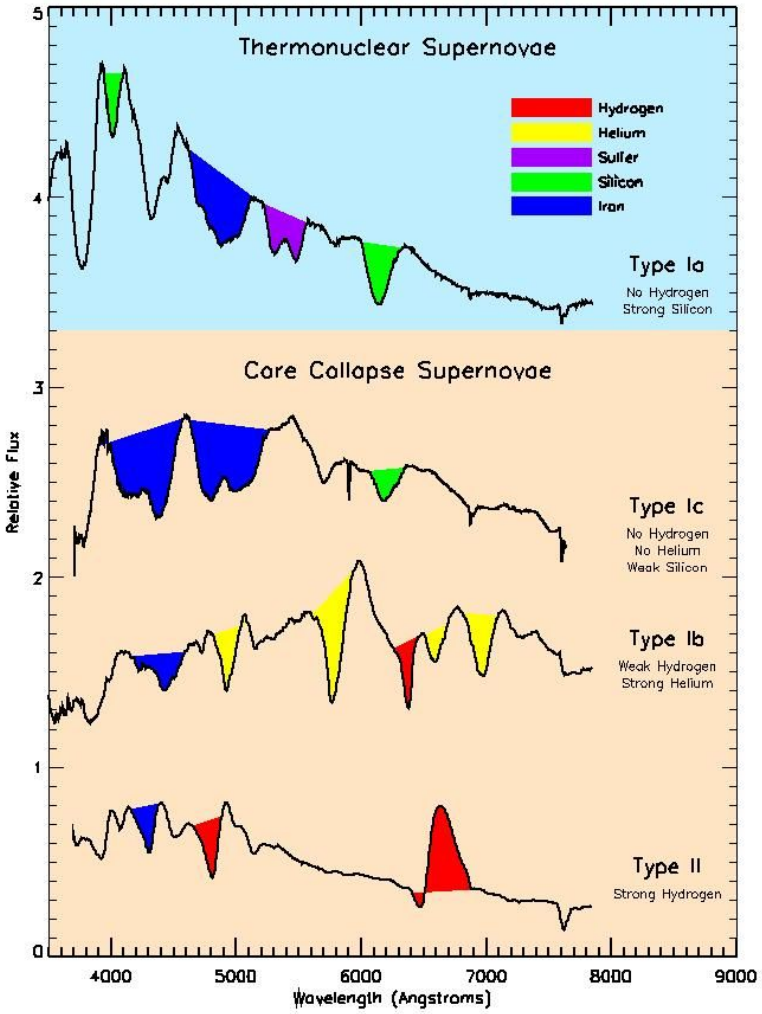
Left - residuals:
Data + signal vs. bkg GP
Data vs. bkg GP

Right:
Extracted signal
→ Final sig kernel hyperpar

Machine Learning @ LPC

Supernova Photometric Classification - LSST





The data Paradigm

year	Number of supernova
1998	42
2014	740
2025	> 10 000

2 million alerts/day
15 TB/day

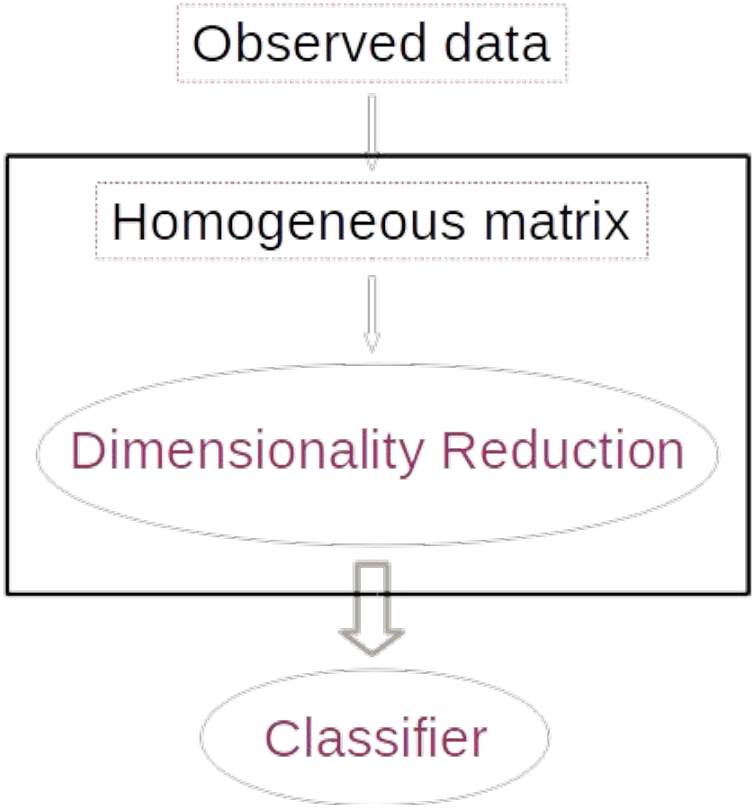
40 nights of LSST

↓

entire Google database

A. Connelly, TED2014

Problem: representativeness



PLAsTiCC

Photometric LSST Astronomical Time-series Classification Challenge

A data challenge aimed to prepare a larger community for the LSST data paradigm

- PI: Renee Hlozek, simulations: Rick Kessler , deployment: Emille Ishida
- SNANA simulations → Light curves in observer-frame (no images!)
- 3 years worth of LSST data, ~ 15 GB
- ~ 3.5 million objects
- A variety of transient models (galactic and extra-galactic, periodic and non-periodic)

Expected release date:

Summer/Fall 2018

- Please respect model-information policy: ``don't ask, don't tell"

- Not all models will be present in the training sample
- Supervised classification + novelty detection
- Deployment: kaggle + RAMP

Slide by Emille Ishida

Machine Learning @ LPC

ATLAS - search new physics over background

Novelty detection

ITT - Infrastructure

High Performanc e Computing

LSST - Transient Classification

PLAsTiCC

Photometric LSST Astronomical Time-series Classification Challenge

PEPS Astro-informatique

TransiXplore

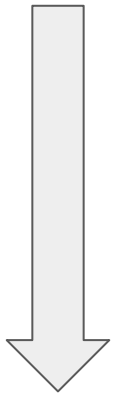
Machine Learning @ LPC

LSST + ATLAS + Sante(?) + Environment(?)

Future Plans

Machine Learning @ LPC

LSST + ATLAS + Sante(?) + Environment(?)



MachineLearning@Cezeaux

Discussions started because of the "Prospective du LPC":

Machine Learning @ LPC

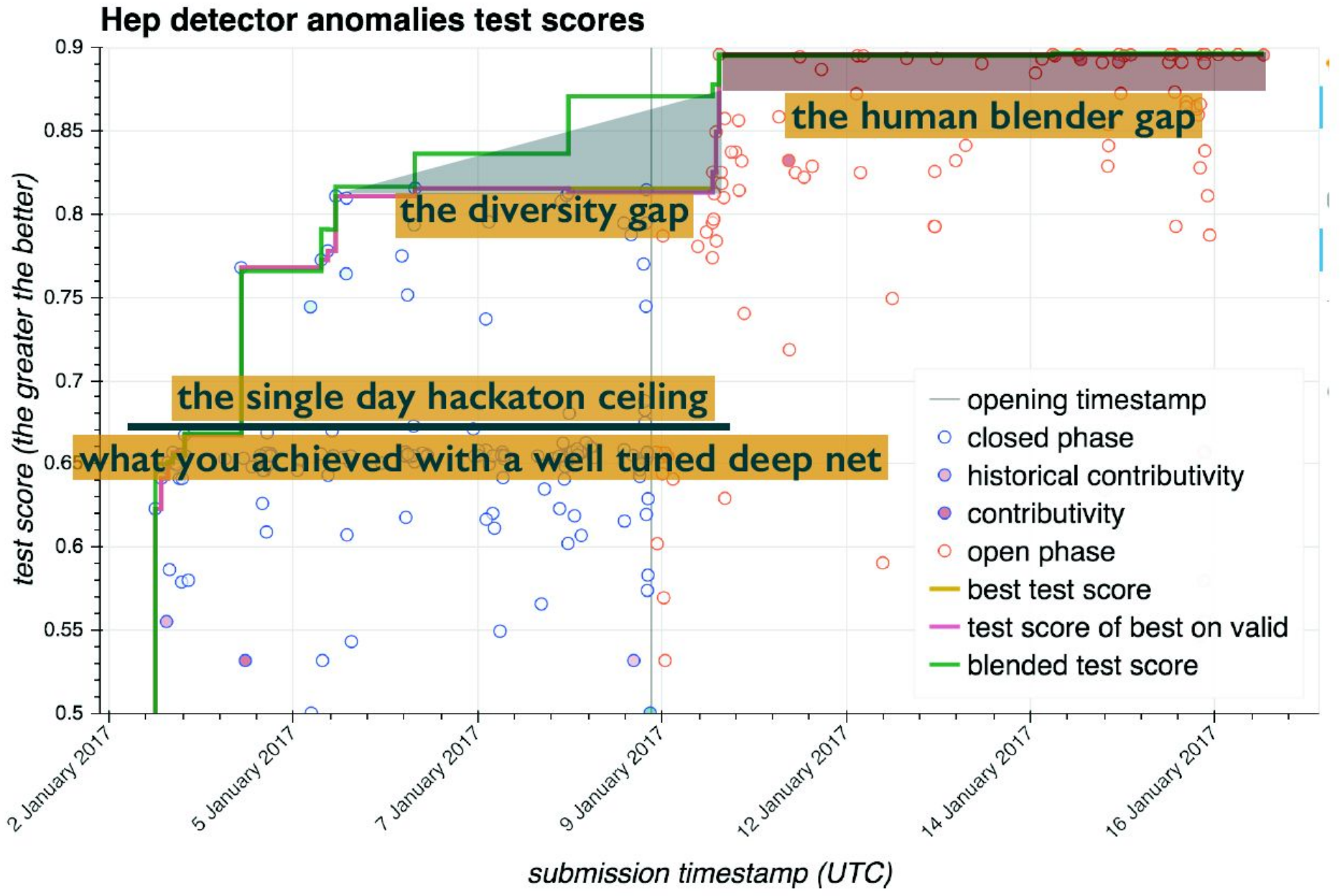
Plan for the near future

Stage 0	Stage 1	Stage 2	Stage 3
2018/1.5	2018/2	2019/1	2019/2
Know thy neighbor	Level the plainfield	Gather tools	The challenge

Optimistic scenario:

We will build - and host - our own data challenge on Autumn/2019 using RAMP

RAMP


Hep detector anomalies test scores
the human blender gap
the diversity gap
the single day hackaton ceiling
what you achieved with a well tuned deep net
opening timestamp
closed phase
historical contributivity
contributivity
open phase
best test score
test score of best on valid
blended test score
test score (the greater the better)
0.9
0.85
0.8
0.75
0.7
0.6
0.6
0.55
0.5
2 January 2017
5 January 2017
7 January 2017
9 January 2017
12 January 2017
14 January 2017
16 January 2017
submission timestamp (UTC)

Main Goals

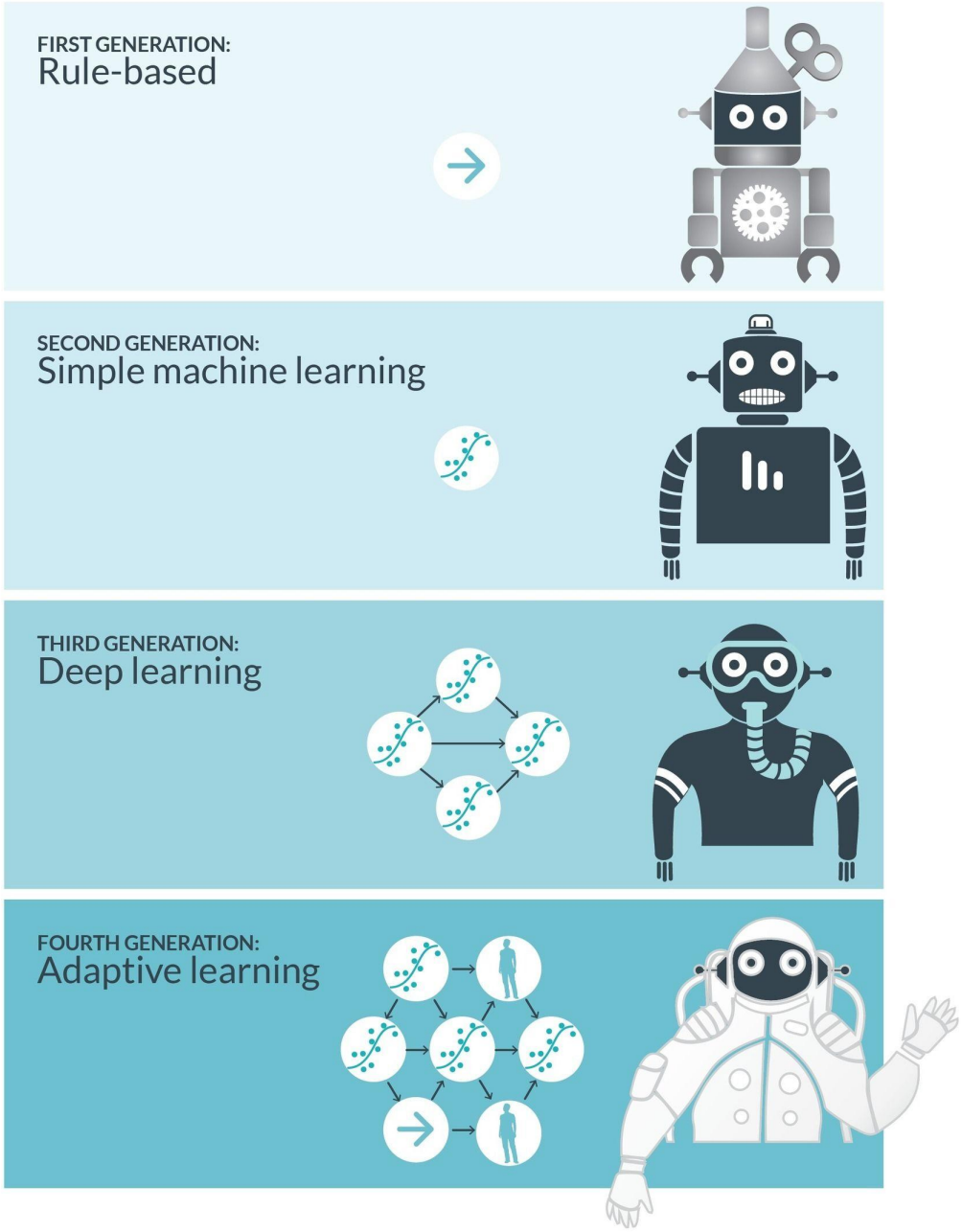
Enabling local interdisciplinary collaboration

- Begin from scratch
  - Build a strong local interdisciplinary community
  - Give students and young researchers practic in transferable skills
  - Build human and material infrastructure for long term collaboration

- Connect with other efforts in France
  - Build strong connection with CDS and INRIA
  - Apply for potential funding to strength collaborations

- Connections with International efforts
  - PLAsTiCC will be an official Kaggle challenge and TransiXplore can provide some resources for those interested
  - This local exercise would also allow the local community to participate in other similar efforts - independent of their domain

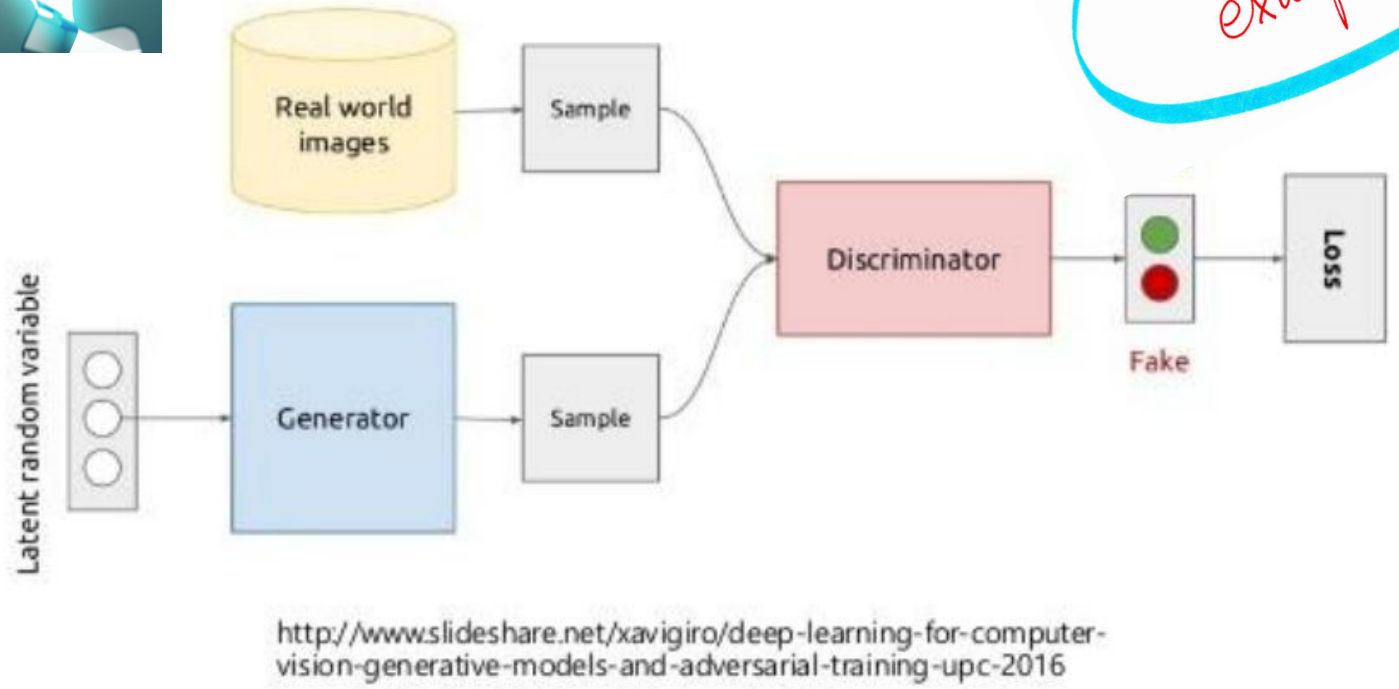
Example



What can be interesting in the future

Adversarial Learning

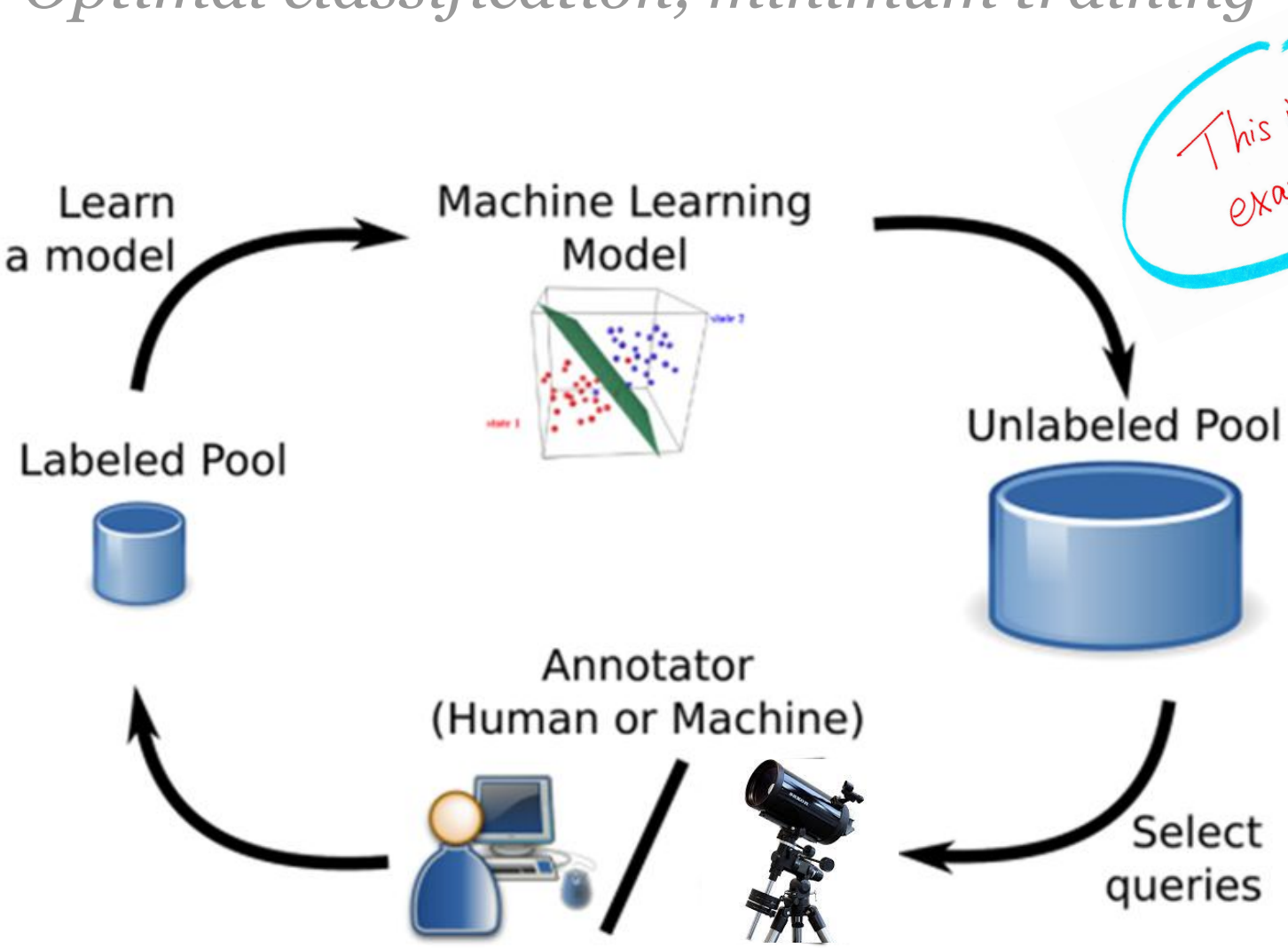
The benefits of a worthy opponent



https://mascherari.press/introduction-to-adversarial-machine-learning/

Active Learning

Optimal classification, minimum training



Summary

There is potential, we need to overcome the barriers!

Extra slides

Active Learning

Optimal classification, minimum training

Can machines learn **better**, with **fewer** labelled examples, if they are carefully chosen?

amazon

35% of Amazon's revenue are generated by it's recommendation engine.

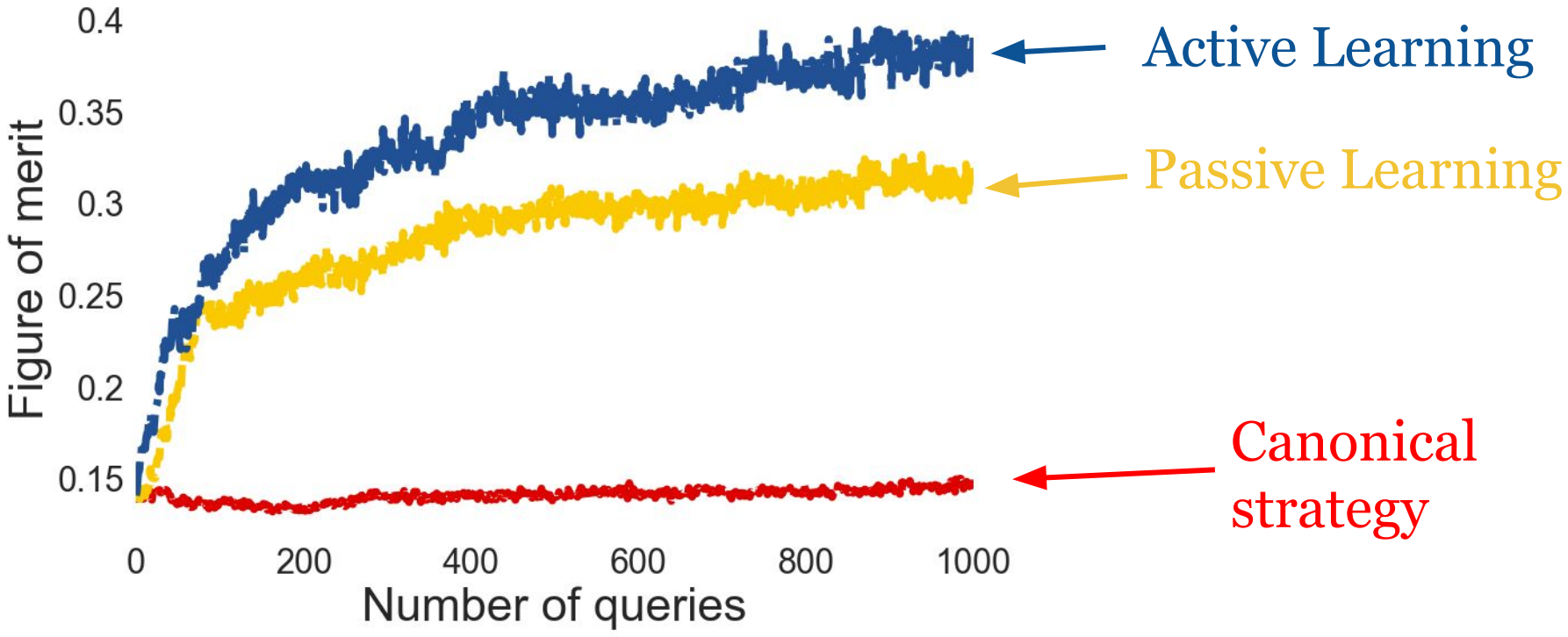
NETFLIX

75% of users select movies based on Netflix's recommendations.

YES!

AL for SN classification

Static results



Ishida** et al., 2018 - arXiv:astro-ph/1804.03765 - from **CRP #4

This is a group effort!

Active Learning result was born in interdisciplinary meeting held in Clermont Ferrand in 2017!

Bring innovation to academia!

Cosmostatistics Initiative

http://cointoolbox.github.io/