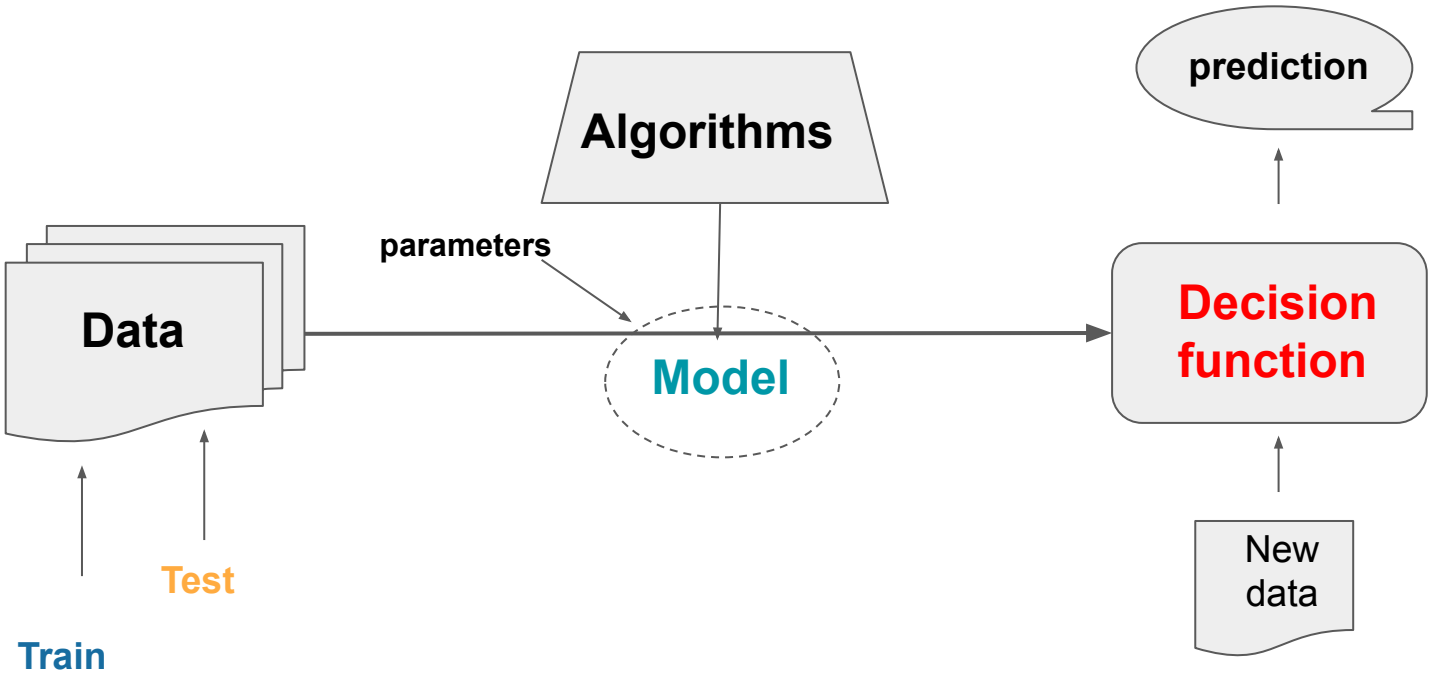
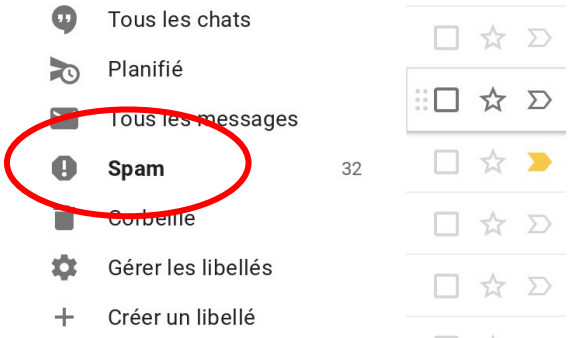

C!RTA ChallenGe 2019
@TIC-HEAP, Constantine
19-21 October 2019
#HEP_OpenData #Python
#DeepLearning #ML
#HiggsData #Visualization
#Coding #SkillSharing
ZiND!
The Abdus Salam International Centre for Theoretical Physics
ICTP
IAEA
RSDT
LPMPS
UNIVERSITÉ DE GENÈVE
FACULTÉ DES SCIENCES
CDTA
cerist
CERN

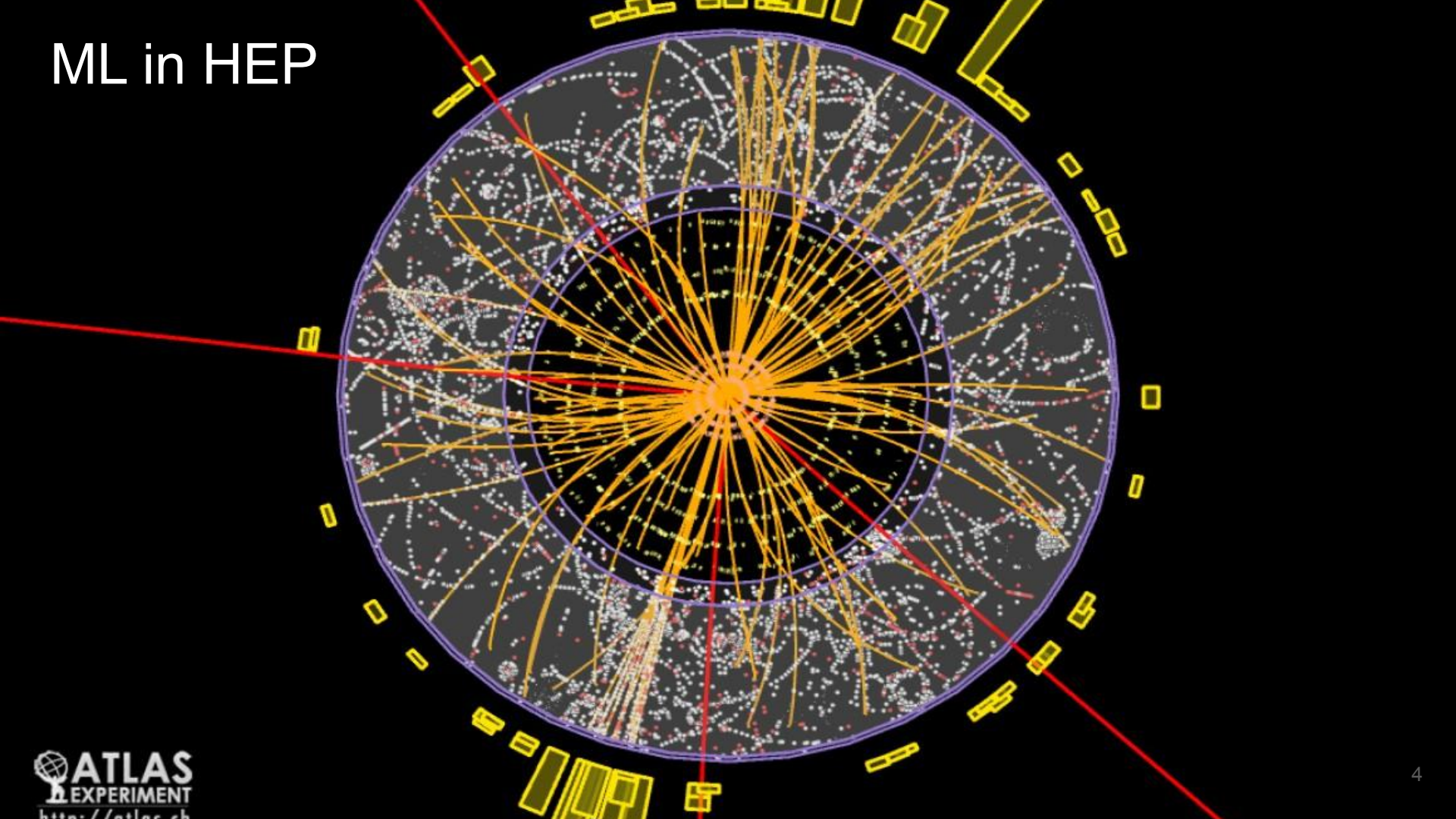
Machine learning in a nutshell



ML in the world: where you already use it



ML in HEP

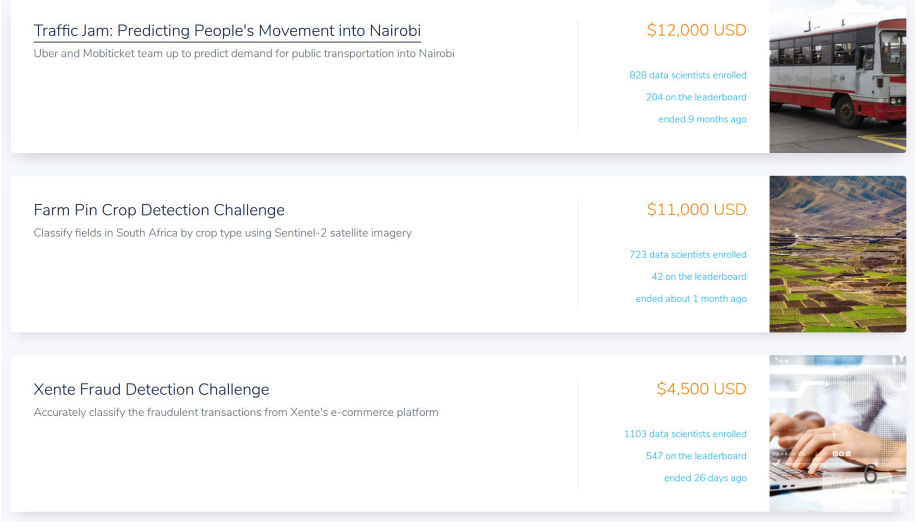


ML in HEP

- Overview here : arXiv:1807.02876
- Most frequently used ML : Boosted Decision Trees (BDTs) and Neural Networks (NN)
- Estimate of a particle’s energy using multiple detectors measurements
- Neural network for merged pixel clusters

(1) Cirta challenge : Particle identification on Zindi

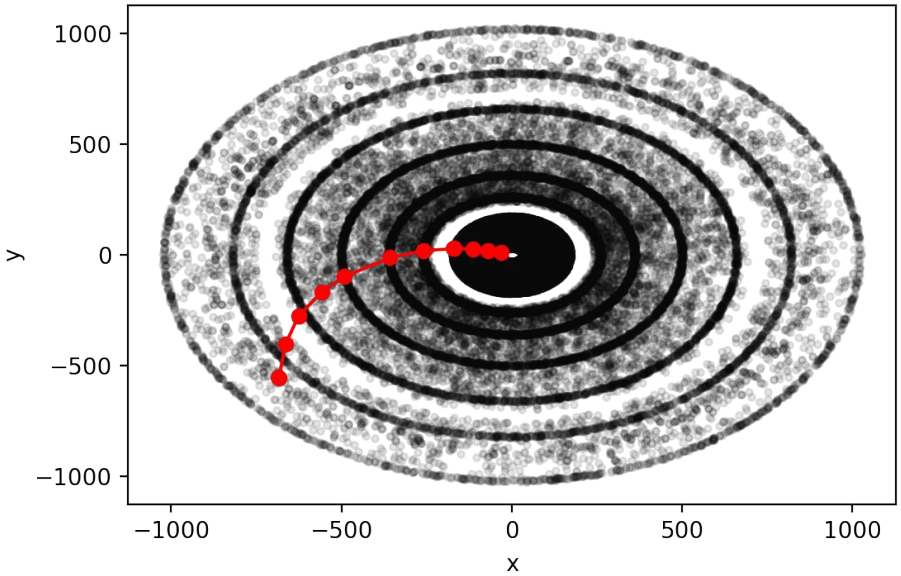
- A community of data scientists solving Africa's toughest challenges
- Proud to add HEP to Africa’s challenges !
- First Algerian challenge as well.
- Find the particle type from its image!

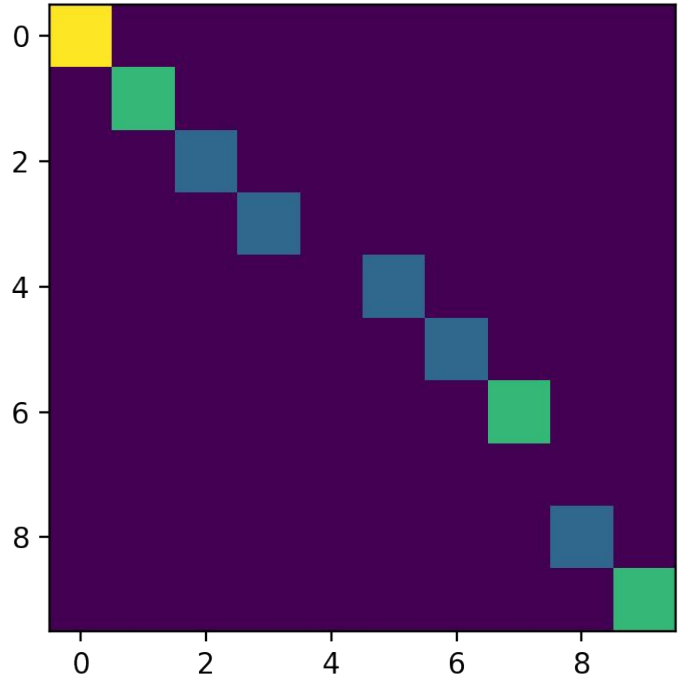


The data : from TrackML challenge

- **Open data simulated for TrackML challenge with ACTS (arXiv: 1904.06778)**
- Labelled particles : tracks + particle_type
- The track is converted to a 2D image
- Learn the type !

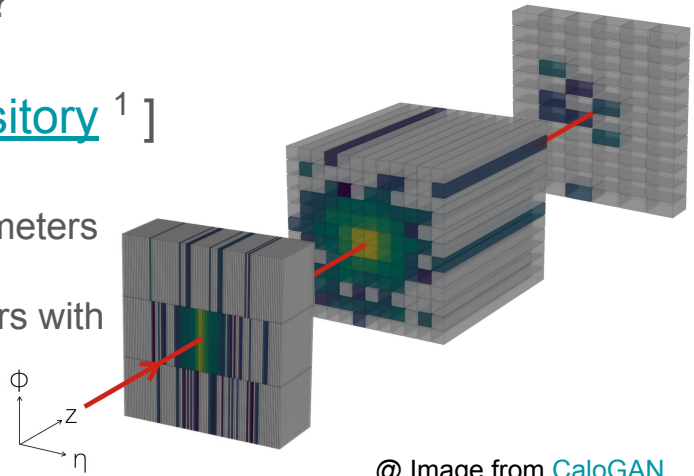
The challenge





(2) Cirta challenge : Particle identification in the calorimeter

- How to classify shower events per particle type?
- Dataset [public dataset @ Mendeley Data repository [1]]
  - Particle Showers in Multi-Layer Electromagnetic Calorimeters
  - The detector is longitudinally segmented into three layers with different granularity
  - The energy depositions of particles passing through the detector can be visualised as a series of 3 images per shower

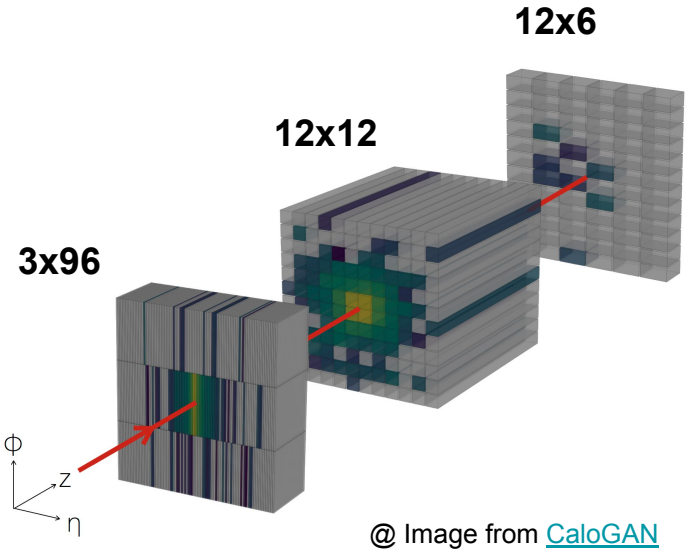


@ Image from CaloGAN

[1] Dataset from : *"Nachman, Benjamin; de Oliveira, Luke; Paganini, Michela (2017), "Electromagnetic Calorimeter Shower Images", Mendeley Data, v1", DOI: 10.17632/pvn3xc3wy5.1*

(2) Cirta challenge : Particle identification in the calorimeter

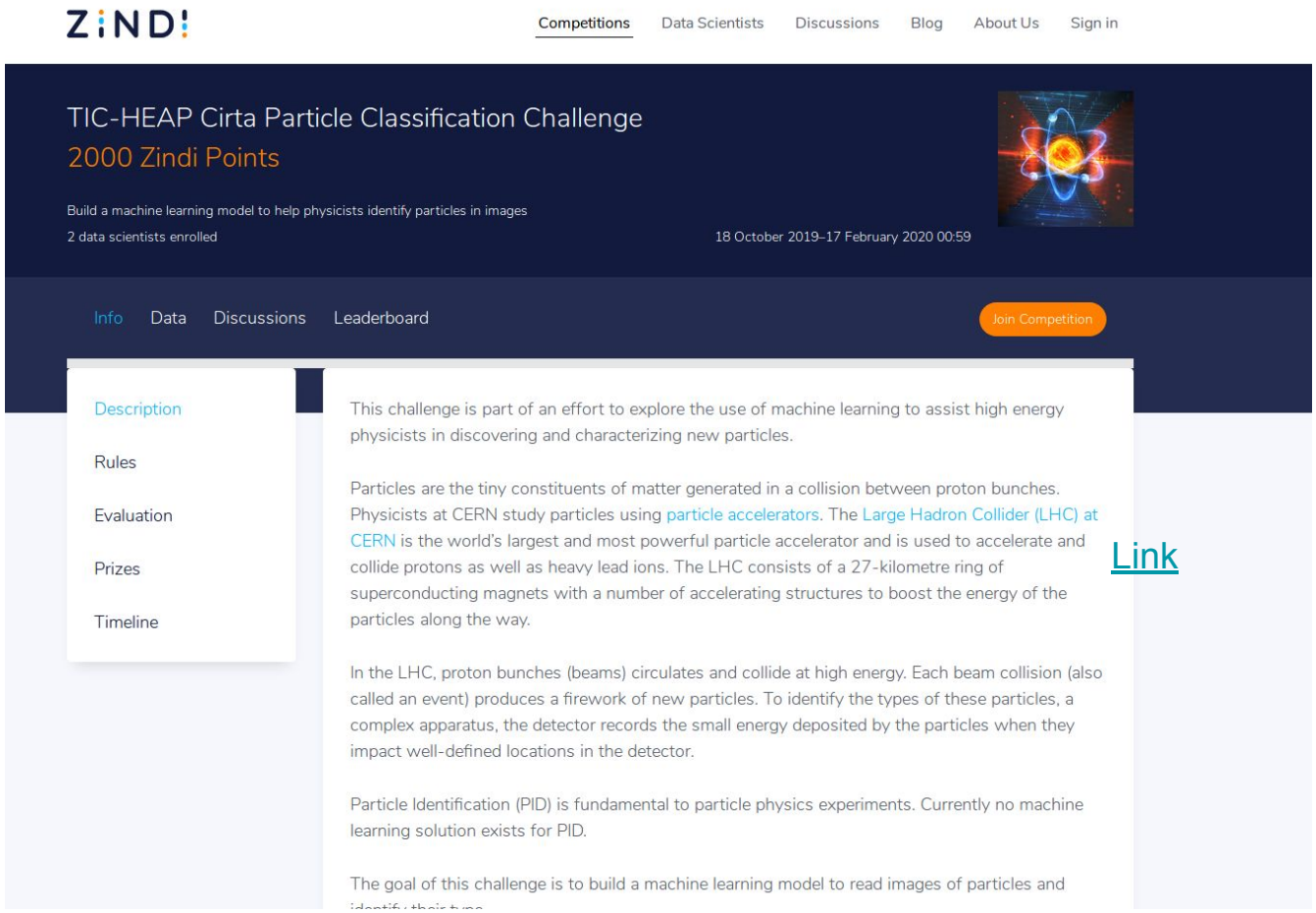
- Dataset
  - 3 files : 1 per particle type: **photons, charged pions, positrons**
  - Each file is a collection of **100.000** calorimeter showers
  - Each file contains the true energy of the incoming particle, the energy per layer (as image data) and an overflow (contains the amount of energy that was deposited outside of the calorimeter section considered)



@ Image from CaloGAN

[1] Dataset from : *"Nachman, Benjamin; de Oliveira, Luke; Paganini, Michela (2017), "Electromagnetic Calorimeter Shower Images", Mendeley Data, v1", DOI: 10.17632/pvn3xc3wy5.1*

zindi.africa/competitions/



Link

CaloCalssification : Github