EASY-ML Hands On Session

Exercise overview

Charles Ndung'u Ndegwa, Isabelle Rocamora, Kinyua Gikunda, Luca Cadamuro

Hands on session

Objective:

- Develop an AI-powered data analysis project using real-world datasets from various scientific domains.

Structure :

- Teams of two will select one of four available project exercises.
- Development will be in Python via Google Colab.

Schedule:

- **Monday, July 21:** Project kick-off and introduction to exercises.
- **Monday, July 21 & Tuesday, July 22:** Full days dedicated to project development.

Support:

- Our staff will be available throughout the two days to provide guidance, answer questions, and offer support.

Particle physics : Higgs boson search with ATLAS

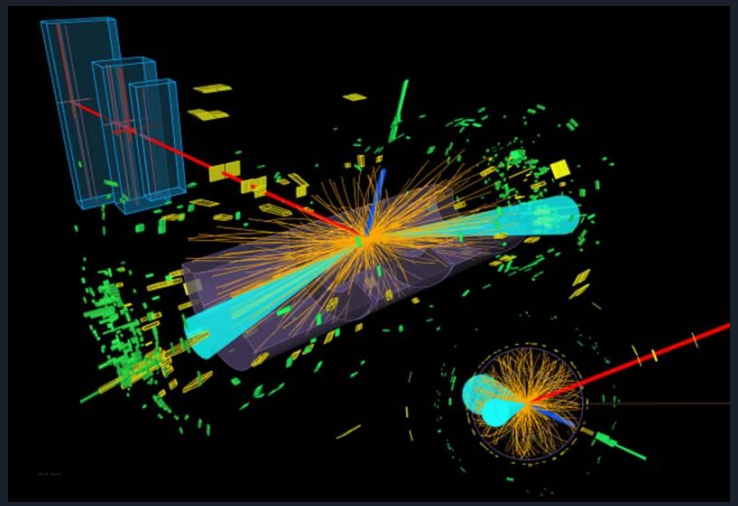
Use data from the ATLAS experiment at the CERN Large Hadron Collider

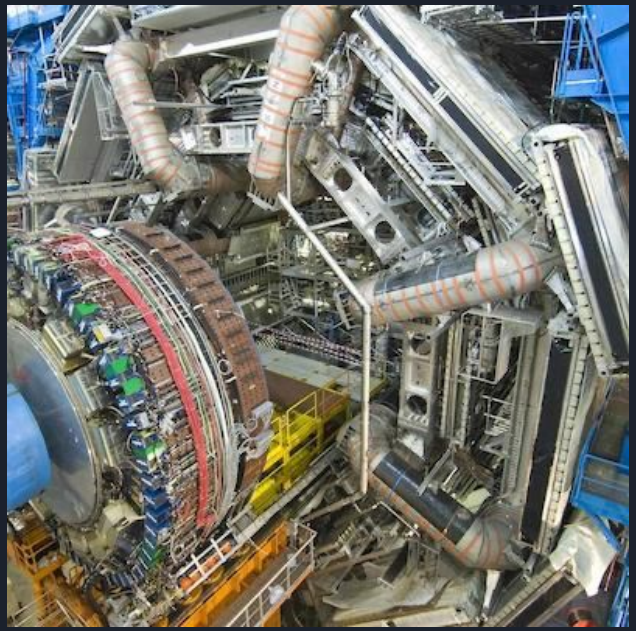
Goal : use AI to search for a very rare Higgs boson signal inside a large background ($H \rightarrow \tau\tau$ decay)

Supervised classification problem

> deploy and optimize AI methods to obtain the best separation between signal and the background

> can be expanded to a regression problem to predict the particle mass from reconstructed data





Geology : flood detection in satellite images

Goal: use satellite images and machine learning to automatically map flooded areas

Case study: April 2024 floods east of Nairobi

Use of Sentinel-2 images : open source high-resolution multispectral images (10 m resolution in the optical and infrared spectrum)

Topics covered :

- Data pre-processing
- Training of supervised classification model (Random Forest and Support Vector Machine)
- Comparison of the performance of different models and the contribution of different spectral information
- Production of a map of flooded areas for visualization on QGIS

Climate : analysis of CO_2 emission data

Objective: Explore CO_2 emission patterns across African countries using unsupervised learning (e.g., K-Means, Autoencoders) using PyTorch.

Dataset: https://www.kaggle.com/datasets/ngaruniki/co2-emission-in-africa

Expected:

1. Jupyter notebook (.ipynb) with code and visualizations.
2. A short report (max 1 pages) explaining findings and insights.

Spectroscopy : Lactose Detection in Milk using NIR Spectroscopy and Machine Learning

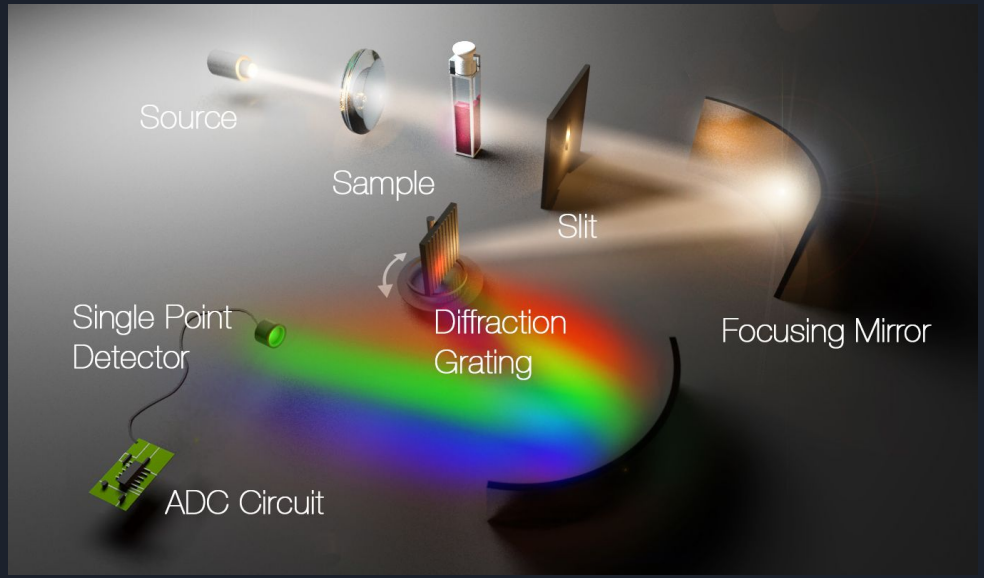
Goal: Identify lactose levels in milk samples using Near Infrared (NIR) spectra.

Workflow Overview:

1. **Dimensionality Reduction:**
   Apply PCA (linear) and Autoencoder (non-linear) to extract latent features.

2. **Classification Models:**
   Train and compare Random Forest, Decision Tree, and 1D CNN classifiers on reduced features.

3. **Evaluation:**
   Use accuracy, F1-score to assess model performance.



Why this matters: Fast, non-destructive food analysis using spectral data and AI!

To conclude

Please fill in your preference for the exercises using this form by the end of this week (July 13th): https://forms.gle/VNiJ15dStL8LAm9r9

> if you have already a partner for the exercise, please indicate it in the form
> we will define teams based on the preferences expressed

Do not hesitate to ask us if you have any question!

Enjoy the school!