The Problem

When there are high luminosity beams, it is important to only write data of interest to disk for analysis. In this project, you will train an ML-based software trigger to classify whether data is "beam" or a "reaction". The data comes in as a point cloud of (x, y, z, q), where z is the beam axis and q is the total charge deposited on the pad associated with the (x, y) coordinate. For this activity, the trace was already processed to provide the z coordinate. There is an additional column of zeros that is unnecessary for this analysis.

You will be building your model using simulated data from a ^{22}Mg alpha scattering experiment that was recently run using the Active-Target Time Projection Chamber at the National Superconducting Cyclotron Laboratory in East Lansing, MI. The data is in the `Mg22_alphaalpha_digiSim.h5` file in this directory. The even events are "reaction" data and the odd events are "beam" data. Your challenge is to take this point cloud data and build a model to classify "reaction" from "beam" events. Note that you will **need** to preprocess this data to put it into a format to use the machine learning methods that we learned about in the summer school (we always have *fixed length* inputs!).

The Write-Up

Your primary deliverable for this assignment is a *Computational Essay* written in `jupyter`. See `https://michellekuchera.com/CSC371/07_project1.html` and `https://hub.gke.mybinder.org/user/mpkuchera-phy200_sp2020-e8h9t5wt/notebooks/Computational_Essay8/Final_Project.ipynb` for inspiration. You will send a link to a GitHub repository that contains the notebook and all other necessary files (e.g. code, images, etc) to Michelle Kuchera and Raghu Ramanujan on the ML Summer School Slack page.

The goal is to communicate what you did in a manner that communicates the work effectively. The purpose of this final project is to demonstrate your ability to step through the workflow of implementing an end-to-end machine learning analysis.

Note that you likely do not want to show *all* your code to the reader in the notebook. You can create user-defined modules and import them in the notebook to hide certain functionality that may not help the reader understand the work.

Here's the overarching writing rule: *you need to be sufficiently precise with your writing and include enough detail that a competent reader could reproduce your results.* Here are some specific things to address in your report, in no particular order, and no matter the format of the report. This is *not* meant to be an exhaustive list.

- What preprocessing did you perform on the data? Describe your exploratory data analysis. Did you generate any plots or charts? Describe these, along with any relevant findings, in your report.
- What models did you build? How do they compare in terms of performance? What was the best performing model, and how did it do?
- What was your model-building and tuning process? How did you address overfitting? How did you make hyperparameter choices?

Recommended Procedure

Here's a recommendation for the steps to take for this project.

- **1:** Explore the dataset, think about feature engineering, build your first model.
- **2:** Run more thorough experiments (hyperparameter tuning, further feature engineering, etc.), analyze your results and iterate, search the literature for related work on the problem, write relevant background information.
- **3:** Complete experiments, take a step back and think about your report's narrative, write drafts, consult with Dr. Kuchera as appropriate.
- **4:** Wrap-up any pending experiments, revise and proof-read the blog post and submit.