# Kaggle competition prompts

This is meant to give you some ideas, roughly in the order of difficulty.

## Level 1

- [ ] Try all four models
- [ ] Can you find better hyperparameters or model architectures than in the notebooks?
- [ ] Visualise the data either as images or data frames. Can you spot any differences between the two categories we are predicting?
- [ ] Can you improve the DNN or BDT by adding or removing some features?
- [ ] Can you make new features for the DNN or BDT that improves the performance?
- [ ] Combine different models into one (e.g. majority voting)
- [ ] Try cross validation - this makes you less likely to drop on the leaderboard when the private leaderboard comes out
- [ ] Play around with creating different graphs for the GNN. Do you need all the nodes or can you ignore some?

## Level 2

- [ ] Can you crop the images or otherwise change them to get a better result
- [ ] Can you create new images with slight variations to the ones we already have to have more data to train on
- [ ] Are the classes perfectly balanced, or does one happen more than another? Can you change the loss function you train to focus more on one of the categories?

## Level 3

- [ ] Are there other variations of the models in the notebooks that could do better on this problem? For instance a Graph Attention Network?
- [ ] Are there other models, e.g. transformers, that could do well on this problem?
- [ ] Can you use pre trained CNNs to get a better performance through transfer learning?
- [ ] Can you find commonalities between the jets that are misclassified? Is there a way to address this?