

CA3109 Practical Session Wednesday Feb 1 - L114, Computing

1. Practical - Part One (Morning session)

Watch: Training a dataset example: <https://www.youtube.com/watch?v=fwY9Qv96DJY>

- a. Get a dataset: (Join Data World)
<https://baryon.be/small-datasets-to-practice-data-visualization-skills/>
<https://data.world/nrippner/titanic-disaster-dataset>
- b. Define your set - X/Y axis (this may already be done)
- c. Open and join <https://jupyter.org/> - or use your own digital notebook
- d. Do a Train Test Split using:
https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html
- e. Decide on a Model - you can use this cheat sheet
https://s3.amazonaws.com/assets.datacamp.com/email/other/ML+Cheat+Sheet_2.pdf
- f. Run a model if you have time, *but note the process.*

2. Part Two (Afternoon session) - group workup of presentations for Friday 3rd

In groups of 4 **choose one area from below**, and find **two papers** that outline ML process in this field. papers should be published no later than 2019.

- Logistics - movement of commercial goods
- Healthcare
- FinTech - digital banking
- AgriTech - crops

Please prepare a presentation on the following: (DL Friday 3rd), 2 students per paper. *Do below points for each paper and then a final comparison of methods.*

1. Brief overview of the area
2. What was the ask?
3. What are the ML methods used,
4. What was the outcome/ results
5. Team : *compare with second paper**

*Ideally we are looking for a comparison of methods used across the two papers chosen.