

Ana's Capstone

Part 1: Capstone Topic + Dataset Validation

THREE(!) ideas

- Predict student's A2 grades and uni entries
- Compare A-level colleges in UK
- Compare different ML methods for text classification

Predict student's A2 grades and uni entries

- **Problem statement:** predict student's A2 grades and uni entry based on past performance.
- **Goals and success criteria:** achieve over 80% accuracy in predictions.
- **Get data from:** school I taught at last year.
- **Potential audience:** school's staff, future students/parents.
- **I love it because:** I'm passionate about education.
- **Problems:** don't have the data (yet); need for anonymity; may not be able to publish all findings; small dataset (one year, 60 leavers).

Compare A-level colleges in UK

- **Problem statement:** show how A-level colleges in the UK compare to each other, based on public results.
- **Goals and success criteria:** have people use my findings (website) when choosing an A-level college. (Not sure if number of hits is a good evaluation measure.)
- **Get data from:** Department of Education public data.
- **Potential audience:** students/parents/teachers looking for an A-level college.
- **I love it because:** I'm passionate about education.
- **Problems:** success criteria is a bit dodgy and hard for me to control; this project would be mostly about visualisations.

Compare different ML methods for text classification

- **Problem statement:** show how different ML methods compare on text classification tasks.
- **Goals and success criteria:** apply new libraries to a well known task; be able to reproduce results in scientific papers.
- **Get data from:** my website.
- **Potential audience:** ML researchers; library users.
- **I love it because:** I have the data (not a very good reason!), can see how much this field evolved in 10 years.
- **Problems:** not too interesting (now); may prove myself wrong.

