Week 7 meeting, datasets, papers and plans

* Overview of this week
  + A proper look at some available data sets and how they will be implemented
  + A list of papers to be used for reference and the first few reports sections
  + Discussion of title ideas and add on options
  + Deliverables ahead of the next meeting
* Datasets
  + Data collection
    - For the introduction/motivational section I think it worth developing a form for gathering information.
    - Looking at people opinion on feedback/marking speeds and quality
    - Unlikely to be using university level essays but good to have the information
    - Speaks to why this project is important
    - Need to look at the ethics approval process.
  + The Hewlett Foundation: Automated Essay Scoring
    - <https://www.kaggle.com/competitions/asap-aes/data>
    - A good primary dataset containing large numbers of essays with varying prompts.
    - Good for ensuring and testing the model’s generalisation capabilities
  + CommonLit Readability Prize
    - <https://www.kaggle.com/competitions/commonlitreadabilityprize/data>
    - Focusing on readability it would be useful for allowing our model to consider more factors. Likely also a useful tool for the feedback generation and/or score explainability
* Papers
  + Automated Essay Scoring: A Survey of the State of the Art
    - <https://arxiv.org/abs/2101.10401>
    - Review of models and includes the hewlett dataset
  + Automated Essay Scoring with Discourse-Aware Neural Models
    - <https://aclanthology.org/W19-4450/>
    - LSTM models
  + Automated essay scoring using efficient transformer-based language models
    - chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://arxiv.org/pdf/2102.13136
  + Discussion boards for both Kaggle datasets
  + Theres lots of work on the two datasets, especially the first. I think that I will do a general lit review for the first submission. And then look at more papers and more specificity in the later versions once I have a clearer plan to the project
* Titles and additional options
  + Handwritten text understanding
  + Feedback generation
  + Explainability to attention diagrams
  + Discussion for titles in the meeting and it can always change, so it’s not that important
* Deliverables
  + Project brief – Oscar
  + Any previous work done by students – Alok (if possible to be shared, good for referencing and understanding the scope that I should be aiming for)