01/11/2019 Meeting 4

Aim: To discuss the chosen datasets, as well as the literature survey and project plan

Project Plan

1. **Reconsider the title**: which do you think is the most suitable?
   Current Title: An analysis of machine learning and deep learning techniques for the detection of sarcasm in text

   Alternatives:
   1. Applying machine learning and deep learning techniques to the detection of sarcasm in text
   2. An analysis of machine learning and deep learning techniques for the detection of sarcasm in text
   3. An analysis of deep-learning based classifiers in the domain of sarcasm detection in text
   4. Detecting sarcasm in text using machine learning and deep learning techniques
   5. A comparative study of machine learning and deep learning approaches to sarcasm detection in text

   2. Have I included enough in the **preliminary preparations** section

2. Do you think I have comprehensively described why this domain is difficult

1. Do you think any terms are missing from the definitions section?

2. In citing a definition of a word, how can I cite a definition taken from one line in an academic paper, especially if I am already citing other areas of the paper

Are you allowed to say e.g. in academic writing
- Yes, it is allowed

Literature Survey: Questions

Theme 2.1: Text Vectorization
- Check the use of technical vocabulary, am I using it in the correct context
- I am having difficulty sourcing the paper that first contains Bag of words and Bag of N-grams, how can I cite this?
- Is the paragraph on Bag of Words and bag of N-grams even relevant?
- I would like to present an alternative paragraph on Word2Vec vs GloVe, do you think this is better? I.e. is the original paragraph too long

- Am I right to italicize new concepts being introduced e.g. *Word2Vec*
- Do the paragraphs look ok, for example the use of new line **and** indent
- How can I cite the dictionary for the first reference

Key Points

- SpaCY is the best library
- Need to fully describe why sarcasm detection is a problem, messy datasets #sarcasm, why is it difficult to detect, comparison with humour, irony etc and the misunderstanding of sarcasm by humans
- Goal is to design a customised approach for this difficult problem, e.g. to highlight the negations / contradictions in sarcasm
- Bert can be used by itself and fine tuned, shows that language models themselves can be fine tuned because they are so powerful

For next week:

- Evaluating datasets (providing summary and stats) - Statistical summary, how many samples, topics etc
- Use topic modelling (Topic modelling for term frequencies (LDA))
- Topics are probability distribution for different words (wordcloud)
- Refer to Noura's email on topic modelling
- Python package called word cloud
- Modify the project plan to include Gantt chart