# 



# Specific Learning Difficulties Cover Note

**Student ID:** 150402149

**Advice for assessors and examiners**

Guidelines for markers assessing coursework and examinations of students diagnosed with Specific Learning Difficulties (SpLDs) –

As far as the learning outcomes for the module allow, examiners are asked to mark exam scripts sympathetically, ignoring the types of errors that students with SpLDs make and to focus on content and the student’s understanding of the subject.

Specific learning difficulties such as Attention Deficit Disorders, dyslexia and or dyspraxia may affect student performance in the following ways:

* The candidate’s spelling, grammar and punctuation may be less accurate than expected
* The candidate’s organisation of ideas may be confused, affecting the overall structure of written work
* The candidate’s proof reading may be weak with some errors undetected, particularly homophones and homonyms which can avoid spell checkers

**Under examination conditions, these difficulties are likely to be exacerbated. Errors are likely to become more marked towards the end of scripts.**

Useful approaches can include:

* Reading the passage quickly for content
* Including positive/constructive comments amongst the feedback so that students can work with specialist study skills tutors on developing new coping strategies
* Using clear English and when correcting; explain what is wrong and give examples
* Using non-red coloured pens for comments/corrections

**Colleagues in schools are asked to ensure that students with specific learning difficulties access the support provided by the** [Disability and Dyslexia Service](http://www.dds.qmul.ac.uk/about/index.html)**.**

For more information regarding marking guidelines see DDS webpage



http://www.dds.qmul.ac.uk/staffinfo/index.html and the [Institutional Marking Practices for Dyslexic Students](http://adshe.org.uk/wp-content/uploads/marking_guidelines_for_good_practice.docx)

**Disability and Dyslexia Service**

Student Services

Room 2.06 Francis Bancroft Building

[www.dds.qmul.ac.uk](http://www.dds.qmul.ac.uk/)

Tel: 020 7882 2756

Email: dds@qmul.ac.uk

### Alteration or misuse of this document will result in disciplinary action

**NLP Project Report**

**Pre-processing**

First, I stripped out the speech marks from the text replacing them with spaces. Next I tokenized using a regex tokenizer that removed all punctuation bar ‘.

Then I expanded out standard contractions that I have compiled in a txt file.

In cases of ambiguity of contractions, I selected one of the pair.

This still leaves some contractions where they are less widely used and more likely to be character specific.

**Improvement**

To make the classifier more advances I tried word2vec. Mostly because I found it interested because of the knowledges that can be represented and explore from the vector relations between words and I wanted to see those results on a more specific data set as in the project.

**Training**

I first trained a word2vector model over the whole of the training data set with the hope that it generalises the data well and will function on the testing set without having trained the model on this set. I used a minimum word frequency of 3 in the word2vector model parameters in a hope to capture as many words in the model as possible because this is a reasonably small dataset.

I created a vector for each sentence by averaging the word2vectors of each word in the given sentences. This should result in a vector that is a descriptor of the sentences semantic meaning.

If a word is not in the word2vector model ignore the word in the averaging.

In addition, I had issues transforming the word2vector model into something that I could classify with. After I had succeeded in transforming the word2vec. I ran the classification and achieved results that were purely random.

The results were effectively random assignment of labels. I decided that this was a dead end to continue working on so moved on to a new method. POS tagging.

**Improvement Attempt 2 POS**

I now augmented my tokens with POS tags and reclassified. The results on my cross validation were poor showing no improvement over the classification without them and no change when stop word prepossessing was removed. To me this implies that my classification method is flawed.

From this I used the nltk named entity tagger to add category labels. I found the tagger was performing poorly and wasn’t identifying names so I removed the processing normalizing the text all to lower case. This resulted in some NE being tagged however these were extremely inaccuracy. Include clear names of people not being identified as people.  
 ‘mickey/NN/B-NP mouse/NN/B-NP’ 🡪 ‘(PERSON Mickey/NNP/O Mouse/NNP/O))’.

**Results**

The results given in the ipython notebook are the result of cross validation as I didn’t managed to progress to the real test data as my results were too unsatisfactory.

**Conclusion**

In conclusion my time management was poor, and my selection of approaches was poor. The result was very slow progress on a word2vec classification method that in the end failed. I should have started with an easier method first and built up rather than jumping to the method that I found most interesting. I ran out of time and was unable to complete a satisfactory project.

I am however, reasonable pleased with my pre-processing.