NLP & ML Assignment 1: Neural PoS tagging

For this assignment, you will create a neural part-of-speech tagger.

The tagger

Implement the tagger in Keras. Use this guide:
  https://nlpforhackers.io/lstm-pos-tagger-keras/
You are welcome to use any neural library for this. Make sure you deal with padding, and that you include an embedding layer.

The data

Download a PoS-tagged corpus from UD (universaldependencies.org). Report which dataset you use. You can select anything you like. What kind of text is it - news, web, conversation, ..?
Make sure you split it into train and test parts. Choose how large to make them, and report this.

The technique

Train a model using the training data, and evaluate over the test data
Report both the token accuracy and complete-sentence accuracy.
Report the accuracy using 10%, 20%, 30%, …. 90% of the training data. Did you have enough data?
Try using some external embeddings.

The analysis

What did the tagger do well? What kind of mistakes did it make? Find what the most common errors were, and give some examples of them - as well as some examples of the tagger working correctly.

Assignment hand-in

Essay
Description of what you did and why, describing your general code, and the answers to all of the above questions in the work description. 500-1000 words.

Code
Include your code. A link to a Colab notebook is best. Test the whole notebook first. I will run it myself.

How?
By Innopolis moodle. I hope it will be working.

When?
The end of September 6