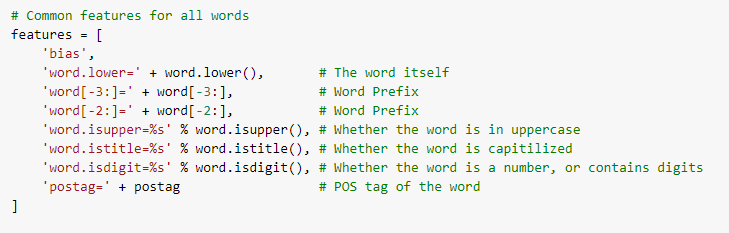
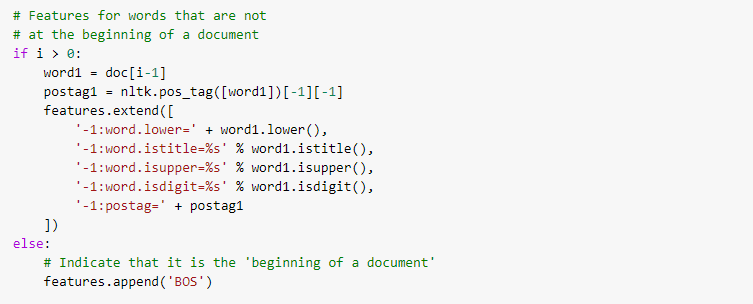
Assignment 3: Named Entity Extraction

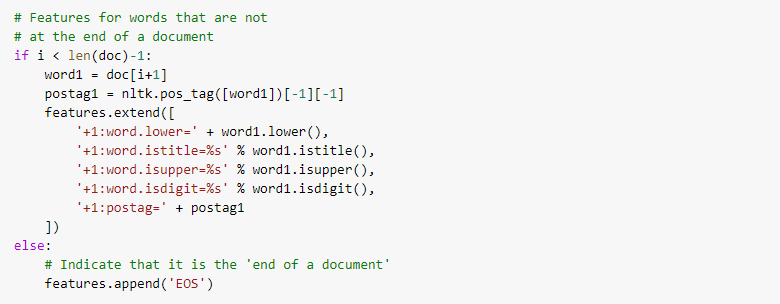
# CRF Model

The list of base features that apply to all words used in the CRF model include:



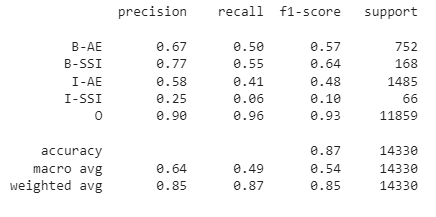
In addition to extracting the features of the given word, we extract features surrounding the word such as previous and next word.



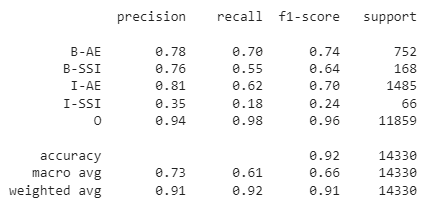


The above features take the context of the word’s surroundings and help with better tagging of I-AE and I-SSI tokens

Baseline Performance:



Improved Performance:



# LSTM Model

Change the LSTM model, by changing its configuration, adding layers, changing the embedding, etc. to further improve the performance from the baseline model provided. You will describe the model changes you made and how these improved the performance in your report.

Baseline

The baseline Bi-LSTM model achieves a macro F1 score of 0.555

Table

Description automatically generated

Number of LSTM layers

Changing the number of LSTM layers to 2, results in a decrease of performance

Table

Description automatically generated

RNN Hidden Dimension

Increasing the dimensions of the hidden vector from 100 to 200 results in a macro F1 score of 0.627

Table

Description automatically generated

# Sequence Labeling

Choose one model and generate the labels in the \_LABELSEQ file format for the test set and submit the file along with your report.

I used the Bi-LSTM model to generate labels on the Test set in the \_LABELSEQ file format and attached it along with this report (test\_labelseq.txt)