Task: Named Entity Recognition

Dataset - [Groningen Meaning Bank](https://www.kaggle.com/abhinavwalia95/entity-annotated-corpus#ner_dataset.csv)

Input: Word, POS

# Target: Tag

Metrics: normalized confusion matrices, precision, recall, F-score (token and entity level)

Models:

* HMM – implement using provided template
* CRF
* Bidirectional LSTM
* BERT
* SpaCy

Scoring:

* EDA: text lengths, vocabulary size, frequent patterns, target tags distribution (+1)
* HMM: fit, predict, smoothing, tupled features (+2)
* CRF: feature engineering, feature importance (+2)
* Bidirectional LSTM: padding, incorporating CNN layers (+2)
* BERT: data preparation, model finetuning (+2)
* SpaCy: model finetuning (+2)
* Evaluation (+1)
* Model efficiency analysis (+1)