

### Approach:

- 1) Dataset was in csv format → Converted Title, Post, and Flair into separate text files.
- 2) Removed punctuation since they do not provide much information – Training becomes faster → Create a dictionary to map words to integers (One hot encoding is necessary to be able to feed into model) → Padded these one hot encoded features to a specific length (which can be decided on the basis of number of words in each title / post, then decide an optimal value) – This is important since inputs must have same length.
- 3) Define a RNN network utilizing embedding and lstm layers and then train.