

In this document I have penned down my approach towards this problem
(4.2 NLP):

NOTE: I could not complete this task. (Didn't have prior experience in NLP unlike CV). Please do go through the attempt.

Task 1,2:

The first step towards this task, was analysing the data and cleaning it. In the dataset there were 238 nan values as 'Posts' and these had to be removed. After this, I decided to merge both the title and posts together as they described the same thing and decided to feed this as the input, labelled with the flairs.

The next step was to find the vector representation of words using GloVe. Next we build RNN/LSTM based models for flair classification task, in which we input the concatenation of title and body of the post and try to predict flair using a softmax classifier.