* My 1st base line approach was using coutvectorizer it got me a public leaderboard score of 0.42.
* Removed digits, emojis, URLs, punctuations, made everything lower
* Later I tried BERT and tuned it for the given data set, removed stop words from the text, as BERT requires more memory if I run on entire passage, I have taken sequence length as 150, but I guess most of the important information is ignored in this approach, here I could not cross the 0.50 on Public LB
* So now I have taken the sentences only in which the given drug is present and used the Bert to classify sentiment which resulted 0.60 on Public LB, on the same I implemented Attentional Encoder Network for Targeted Sentiment Classification which resulted in 0.56 on Public LB.
* My final model is ensemble of 3 BERT and 1 AEN, Loss function that I used is CrossEntropyLoss with class weights = 1/number of observation in each corresponding class
* My key take away will be to try different things and check what works according to the data. spend some time on listing what are all the things to try.