Aarya Patel – Contest 5

So my first strategy for this challenge was using a simple pipeline model by sklearn which was a count vectorizer, tdidf transformer, and a multinomial naive bayes classifier put together in a single object without a lot of setup. It was quick to implement after figuring out how to process the input data, but it didn't really work well in terms of accuracy. So my final strategy used was running the data set on a pre-trained glove to create word-embeddings with a vocabulary size of 20000. But before this, I tokenized the data set using a keras tokenizer to make the many sequences of words which was used to create the word_embeddings which a model (did a bit of research to find a good one which incorporated Conv1D layer, maxpooling1d, and LSTM) ran through.