(see table below for parts a and b)

Model	max_iter	Accuracy	Log Likelihood
NB	N/A	0.98	-504051
NB+EM	1	1	-849692
NB+EM	2	0.81	-848118
NB+EM	10	0.69	-846701

c. I believe that the NaiveBayesEM models have a lower likelihood than the NaiveBayes model because the log likelihood of the NaiveBayesEM model takes into account the all of the pre-1964 speeches. The likelihood of seeing a specific set of words in all speeches *including* those prior to 1964 is necessarily going to be lower than the likelihood of seeing a specific set of words in a smaller subset of those speeches. This behavior is analogous to the probability of guessing the outcomes of a sequence of coin flips; the odds of guessing the sequence correctly go down as you increase the number of flips. The accuracy of the 1-iteration NB+EM and NB are close to 1, which makes sense, as our alpha and beta parameters are fit to this data knowing how these speeches are classified – they are therefore able to predict them when tested. As we add more iterations to our EM algorithm, however, the accuracy goes down even as the log likelihood increases. This is likely because the speeches before 1964 do not follow the same word to party relationships as those after 1964. By including this extra data, our model is beginning to classify the post-1964 speeches as belonging to clusters that do not align as well with modern-day partisan politics.