Remember to split data into training and testing – you should mention this in exam questions unless otherwise specified.

Include both the formula for Naïve Bayes in the question (derivation is not necessary) – P(c) and p(w_i|c). Also mention how you would evaluate your system? Run it on the test data and calculate the accuracy, or precision or recall etc.

Remember to give formulae!

More annotations makes the sentiment closer to the ground truth. Annotation is always very subjective – especially in this example (bullying). The reason to get more annotations is to make the "average score over a larger number of people is closer to ground truth".

Learn the formula for κ.

If you have multiple annotations then you should propose a mapping from the annotations to the ground truth.

"Tuning Naïve Bayes" is only changing the parameter and the cutoff point.

The only way of evaluating (how much a system has grown out of date) is to be evaluated. Again, re-use parts of earlier questions in later questions.

If you say "test it on data" – state where you get the data from! Getting data is one of the largest problems in machine learning.

In language change there are two things which are relevant

- New words

- Old words becoming a threat (Wayne Rooney Effect)