7 Machine Learning and Real-world Data (fm611)

You are tasked with developing a detector of offensive messages using a text classification method. For this purpose a dataset has been provided to you. The dataset contains a sample of 10,000 messages which have been labelled for offensiveness. The files contain information on the text of a message, together with a label ("OK"/"offensive")

(*a*) You want to build your detector using an 'offensiveness' lexicon.

(*i*) Which steps would you follow to develop such a lexicon? [2 marks]

(*ii*) How would the detector work? [3 marks]

(*iii*) How would you evaluate the performance of the detector? Give details of the evaluation metrics and equations, as well as necessary data splits. [4 marks]

(*b*) You have the budget to employ 10 human raters to rate 100 texts each. You want to use these ratings to evaluate your system. How would you assess whether the resulting human ratings are reliable for evaluating your system? [3 marks]

(*c*) Explain how you can set up a Naive Bayesian classifier for this task and derive the required parameter estimates.

(*i*) Give all necessary formulae and discuss smoothing if necessary. [4 marks]

(*ii*) In what circumstances is the Naive Bayes classifier an improvement over the lexicon-based approach? Explain your answer fully and discuss how we might practically use the lexicon to improve the Naive Bayes approach further. [4 marks]