Intro to AI coursework – PLAN

QUESTION 1:

You should consider the following aspects:

* The kind of algorithm to use (e.g.: classification/regression/clustering)
* The metric to use to measure the performance of the model
* What sort of baseline to compare the model to (sklearn has a module sklearn.dummy which may be useful in generating a baseline)
* How to choose the hyperparameters of your model
* How to test the performance of your model

Concretely, you should use two algorithms from scikit-learn and compare their performance on the dataset. You should also compare the performance of your chosen models against a baseline–i.e. a simple model that more complex models should be able to beat. sklearn has a module sklearn.dummy which may be useful in generating a baseline. You should use techniques to assess the ability of the models to generalise to unseen data and to ensure that your assessment of the models’ performance is robust.

Material from worksheets 13, 14, 16, and 17 will be helpful here.

Your answer to this question should take the form of a short report (maximum 4 pages), together with commented code, detailing the approach you will take. Make sure you address all the bullet points above, and explain your decisions. For example: ‘I chose to use a X algorithm because Y’. ‘Because of Z, I used metric M’. You should use plots and figures as appropriate to illustrate your decisions.

The code will not be marked for elegance, but it should run correctly. If you are using jupyter, a good tip is to make sure you have restarted the kernel and made sure that the code can run from scratch before submitting.