IT3212 Assignment 2

Main Assignment (Minimum marks to get the assignment approved 65)

Now you have formulated your problem statement in both worlds (real world and machine learning). So it's time to solve the problem. To solve the problem, you will need to complete the following steps:

1. implement the preprocessing (**15%) → you don't have to justify this because this was part of the first assignment**
2. extract features (**15%, without justification 5%**)
3. select features (**15%, without justification 5%**)
4. implement two basic modeling methods based on the selected features (**10%, without justification 5%**), with hyperparameter tuning (**5%**), explain the reason for choosing specific hyperparameters to tune (**10%**)
5. select two performance metrics and compare the modelling methods (**5%**). Explain the reasons why one is better than the other, or why they perform similar (**15%**)

At all the steps of the pipeline above, you need to justify the choice of all the methods used.

Apart from that, design (not develop) your pipeline based on one Advanced modelling techniques. Again, justify your choices (**10%**).

Include individual contributions at the end of the submission (This is mandatory section but there are no marks for this).

The submission should be between 1500 and 3000 words.

Personalized feedback form

This is the part where you can choose to have me focus on some/all the aspects of your submission. Remember the following points: 1. You have to submit this form as a group, so have a discussion among the members before submitting this form. 2. You can choose all the given options. However, please try to reflect and carefully select the ones that you want me to focus on. You will also receive a general and in-depth feedback by default.

Feedback options:

Choose between zero and all of the following aspects of the assignment for me to focus my feedback on (delete the option that you do not want).

1. Feature extraction (YES/NO)
2. Feature selection (YES/NO)
3. Choice of basic modelling methods (YES/NO)
4. Choice of performance metrics (YES/NO)
5. Comparison of modelling methods (YES/NO)
6. Advanced Pipeline Design (YES/NO)