

WQD 7006: Machine Learning for Data Science
Quiz 2020

20 Minutes/5 marks

Please save your solution using WORD document, and save the file using your full name (No ID etc. please). Submit via Spectrum.

1. The dataset below is used to learn a decision tree which predicts if people pass machine learning (True or False), based on their previous GPA (High, Medium, or Low) and whether or not they studied. Use log base 2 for all your calculations. Show the steps.

GPA	Studied	Passed
L	F	F
L	T	T
M	F	F
M	T	T
H	F	T
H	T	T

- i. What is the entropy $H(\text{Passed})$? (1 mark)
- ii. What is the entropy $H(\text{Passed} \mid \text{GPA})$? (1 mark)
- iii. What is the entropy $H(\text{Passed} \mid \text{Studied})$? (2 marks)
- iv. Draw the full decision tree that would be learned for this dataset. (1 mark)