2. Table 1 below contains students exam performance data for six students, listing details such as if they achieved an 'A' last year (yes/no), their gender (male/female), did they have good attendance (yes/no), did they spend too much time partying (yes/no) and did they achieve A this year (yes, no).

No.	Student	A last year	Gender	Attendence	Partying	A this year
1	Mick	yes	male	no	yes	yes
2	Alan	yes	male	yes	no	yes
3	Ellie	no	female	yes	no	yes
4	Jack	no	male	no	yes	no
5	George	yes	female	yes	yes	yes
6	Simon	no	male	yes	yes	no

Table 1: Dataset showing exam performance of a set of six students.

Table 3 at the end of this exam paper contains equations you may find useful when answering this question.

(a) Using the dataset in Table 1, what is the entropy of this set of training examples with respect to the target feature classification?

(5 marks)

(b) Construct the decision tree that would be generated by the **ID3** algorithm using entropy-based information gain. Show the steps when building the tree.

(20 marks)

$$H(t, \mathcal{D}) = -\sum_{l \in levels(t)} \left(P(t = l) \times log_2(P(t = l)) \right)$$

$$rem(d, \mathcal{D}) = \sum_{l \in levels(d)} \underbrace{\frac{|\mathcal{D}_{d=l}|}{|\mathcal{D}|}}_{weighting} \times \underbrace{H(t, \mathcal{D}_{d=l})}_{\substack{entropy\ of \\ partition\ \mathcal{D}_{d=l}}}$$

$$IG(d, \mathcal{D}) = H(t, \mathcal{D}) - rem(d, \mathcal{D})$$