2. A convicted criminal who reoffends after release is known as a *recidivist*. The table below lists a dataset that describes prisoners released on parole, and whether they reoffended within two years of release.[1]

ID	Good Behavior	Age < 30	Drug Dependent	Recidivist
1	false	true	false	true
2	false	false	false	false
3	false	true	false	true
4	true	false	false	false
5	true	false	true	true
6	true	false	false	false

This dataset lists six instances where prisoners were granted parole. Each of these instances are described in terms of three binary descriptive features (Good Behavior, Age < 30, Drug Dependent) and a binary target feature, Recidivist. The Good Behavior feature has a value of *true* if the prisoner had not committed any infringements during incarceration, the Age < 30 has a value of *true* if the prisoner was under 30 years of age when granted parole, and the Drug Dependent feature is *true* if the prisoner had a drug addiction at the time of parole. The target feature, Recidivist, has a *true* value if the prisoner was arrested within two years of being released; otherwise it has a value of *false*.

a. Using this dataset, construct the decision tree that would be generated by the **ID3** algorithm, using entropy-based information gain.

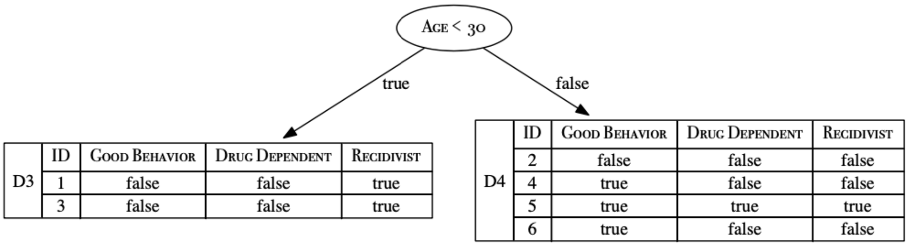
The first step in building the decision tree is to figure out which of the three descriptive features is the best one on which to split the dataset at the root node (i.e., which descriptive feature has the highest information gain). The total entropy for this dataset is computed as follows:

$$
\begin{aligned}
& H\left(\textsc{Recidivist}, \mathcal{D}\right) \\
&= - \sum_{l \in \left\{ \substack{\textit{true}, \\ \textit{false}} \right\}} P(\textsc{Recidivist} = l) \times log_2 \left(P(\textsc{Recidivist} = l)\right) \\
&= - \left(\left({}^3/_6 \times log_2({}^3/_6) \right) + \left({}^3/_6 \times log_2({}^3/_6) \right) \right) = 1.00 \; \textit{bit}
\end{aligned}
$$

The table below illustrates the computation of the information gain for each of the descriptive features:

Split by Feature	Level	Part.	Instances	Partition Entropy	Rem.	Info. Gain
GOOD BEHAVIOR	*true*	$\mathcal{D}_1$	$\mathbf{d}_4, \mathbf{d}_5, \mathbf{d}_6$	0.9183	0.9183	0.0817
	false	$\mathcal{D}_2$	$\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3$	0.9183		
AGE < 30	*true*	$\mathcal{D}_3$	$\mathbf{d}_1, \mathbf{d}_3$	0	0.5409	0.4591
	false	$\mathcal{D}_4$	$\mathbf{d}_2, \mathbf{d}_4, \mathbf{d}_5, \mathbf{d}_6$	0.8113		
DRUG DEPENDENT	*true*	$\mathcal{D}_5$	$\mathbf{d}_5$	0	0.8091	0.1909
	false	$\mathcal{D}_6$	$\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3, \mathbf{d}_4, \mathbf{d}_6$	0.9709		

AGE < 30 has the largest information gain of the three features. Consequently, this feature will be used at the root node of the tree. The figure below illustrates the state of the tree after we have created the root node and split the data based on AGE < 30.



	ID	GOOD BEHAVIOR	DRUG DEPENDENT	RECIDIVIST
D3	1	false	false	true
	3	false	false	true

	ID	GOOD BEHAVIOR	DRUG DEPENDENT	RECIDIVIST
D4	2	false	false	false
	4	true	false	false
	5	true	true	true
	6	true	false	false

In this image we have shown how the data moves down the tree based on the split on the AGE < 30 feature. Note that this feature no longer appears in these datasets because we cannot split on it again.

The dataset on the left branch contains only instances where RECIDIVIST is *true* and so does not need to be split any further.

The dataset on the right branch of the tree ($\mathcal{D}_4$) is not homogenous, so we need to grow this branch of the tree. The entropy for this dataset, $\mathcal{D}_4$, is calculated as follows:

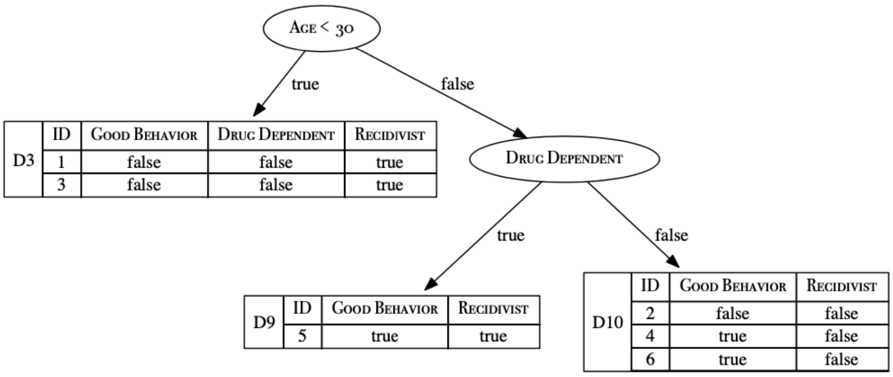
$$
\begin{aligned}
&H(\text{RECIDIVIST}, \mathcal{D}_4) \\
&= -\sum_{l \in \left\{\substack{true,\\ false}\right\}} P(\text{RECIDIVIST} = l) \times log_2\left(P(\text{RECIDIVIST} = l)\right) \\
&= -\left(\left({}^1\!/_4 \times log_2({}^1\!/_4)\right) + \left({}^3\!/_4 \times log_2({}^3\!/_4)\right)\right) = 0.8113\ bits
\end{aligned}
$$

The table below shows the computation of the information gain for the GOOD BEHAVIOR and DRUG DEPENDENT features in the context of the $\mathcal{D}_4$ dataset:

Split by Feature	Level	Part.	Instances	Partition Entropy	Rem.	Info. Gain
GOOD BEHAVIOR	*true*	$\mathcal{D}_7$	$\mathbf{d}_4, \mathbf{d}_5, \mathbf{d}_6$	0.918295834	0.4591	0.3522
	false	$\mathcal{D}_8$	$\mathbf{d}_2$	0		
DRUG DEPENDENT	*true*	$\mathcal{D}_9$	$\mathbf{d}_5$	0	0	0.8113
	false	$\mathcal{D}_{10}$	$\mathbf{d}_2, \mathbf{d}_4, \mathbf{d}_6$	0		

These calculations show that the DRUG DEPENDENT feature has a higher information gain than GOOD BEHAVIOR: 0.8113 versus 0.3522 and so should be chosen for the next split.

The image below shows the state of the decision tree after the $\mathcal{D}_4$ partition has been split based on the feature DRUG DEPENDENT.



D3

ID	GOOD BEHAVIOR	DRUG DEPENDENT	RECIDIVIST
1	false	false	true
3	false	false	true

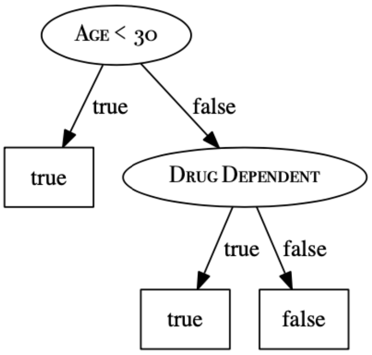
D9

ID	GOOD BEHAVIOR	RECIDIVIST
5	true	true

D10

ID	GOOD BEHAVIOR	RECIDIVIST
2	false	false
4	true	false
6	true	false

All the datasets at the leaf nodes are now pure, so the algorithm will stop growing the tree. The image below shows the tree that will be returned by the ID3 algorithm:



b. What prediction will the decision tree generated in part (a) of this question return for the following query?

$$\text{GOOD BEHAVIOR} = \textit{false}, \text{AGE} < 30 = \textit{false},$$
$$\text{DRUG DEPENDENT} = \textit{true}$$

RECIDIVIST = *true*

c. What prediction will the decision tree generated in part (a) of this question return for the following query?

$$\text{GOOD BEHAVIOR} = \textit{true}, \text{AGE} < 30 = \textit{true},$$
$$\text{DRUG DEPENDENT} = \textit{false}$$

RECIDIVIST = *true*