CS 422 Quiz 2 Name _____

1. Build the best decision tree using the following data. Use the Information Gain algorithm from class.

x_1	x_2	label
0	0	0
0	1	1
1	0	1
1	1	1

Table 1: Training data.

$p(c)$	$p(c)log_2(p(c))$
$\frac{1}{4}$	-0.5
$\frac{1}{2}$	-0.5
$\frac{3}{4}$	-0.31
1	0

Table 2: Log table.

CS 622 Quiz 2 Name _____

1. Show why Information Gain is a better measure of success than accuracy when building a decision tree for the following data.

x_1	x_2	label
0	0	0
0	1	1
1	0	1
1	1	1

Table 3: Training data.

$$IG = H - \sum_{t \in T} p(t)H(t)$$

$$H = \sum_{c \in C} -p(c)log_2(p(c))$$