Part 3

Question 1:

Let p(T) be the probability of the positive class
Let p(F) be the probability of the negative class

Entropy(S) = $-p(T)\log_2(p(T)) - p(F)\log_2(p(F))$

$p(T) = 6/10 = 0.6$
$p(F) = 4/10 = 0.4$

Entropy(S) = $-(0.6\log_2(0.6) - 0.4\log_2(0.4)$
$\log_2(0.6) \approx -0.737$, $\log_2(0.4) \approx -1.322$
Entropy(S) $\approx$ $-(0.6(-0.737) + 0.4(-1.322))$
$\approx 0.4422 + 0.5288$
≈ 0.971

Therefore, the entropy of this collection of training examples with respect to the target class is approximately 0.971.

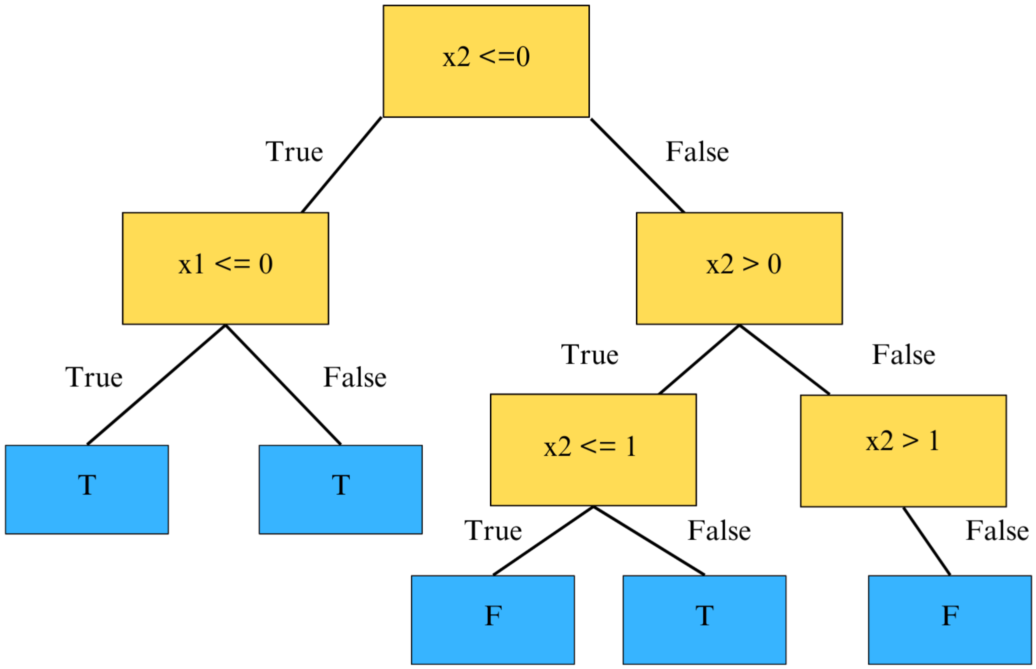
Question 2:

There are two attributes as options for the first split, x1 and x2. x1 is a binary attribute that can take values of 0 or 1. A first split on x1 would create two branches, one branch for x1 = 0 and another for x1 = 1. x2 is an ordinal categorical attribute that takes values of 0, 1, and 2. For this binary decision tree, we handle the three values of x2 by making two binary splits. As the first split, we can split on x2 <= 0 giving one branch for x2 = 0 and another branch for x2 > 0. For the second split in this tree, we handle the remaining values x2 = 1 and x2 = 2 by splitting on x2 <= 1. This gives us one branch for x2 = 1 and another branch for x2 = 2. While x2 has three distinct values, we handle it with two binary splits, ensuring that the tree remains binary.

Question 3:

Information gain for split on x1: 0.02
Information gain for split on x2: 0.247

Question 4:


x2 <=0
True
False
x1 <= 0
x2 > 0
True
False
True
False
T
T
x2 <= 1
x2 > 1
True
False
False
F
T
F