1. Entropy of Y

   S = [12+,9-] → P(c1) = 12/21 & P(c2) = 9/21

   H(Y) = -(12/21)log(12/21)-(9/21)log(9/21) = **.985**

2. Information Gain

   Values for (x1) = T,F

   S|T → [7+,1-] → P(c1) = 7/8 & P(c2) = 1/8
   S|F → [5+,8-] → P(c1) = 5/13 & P(c2) = 8/13

   Entropy of T → -(7/8)log(7/8)-(1/8)log(1/8) → .543
   Entropy of F → -(5/13)log(5/13)-(8/13)log(8/13) → .961

   IG(x1) = H(Y) - H(Y|x1) = .98522 - (8/21)(.543) - (13/21)(.961) = **.183**

   Values for (x2) = T,F

   S|T → [7+,3-] → P(c1) = 7/10 & P(c2) = 3/10
   S|F → [5+,6-] → P(c1) = 5/11 & P(c2) = 6/11

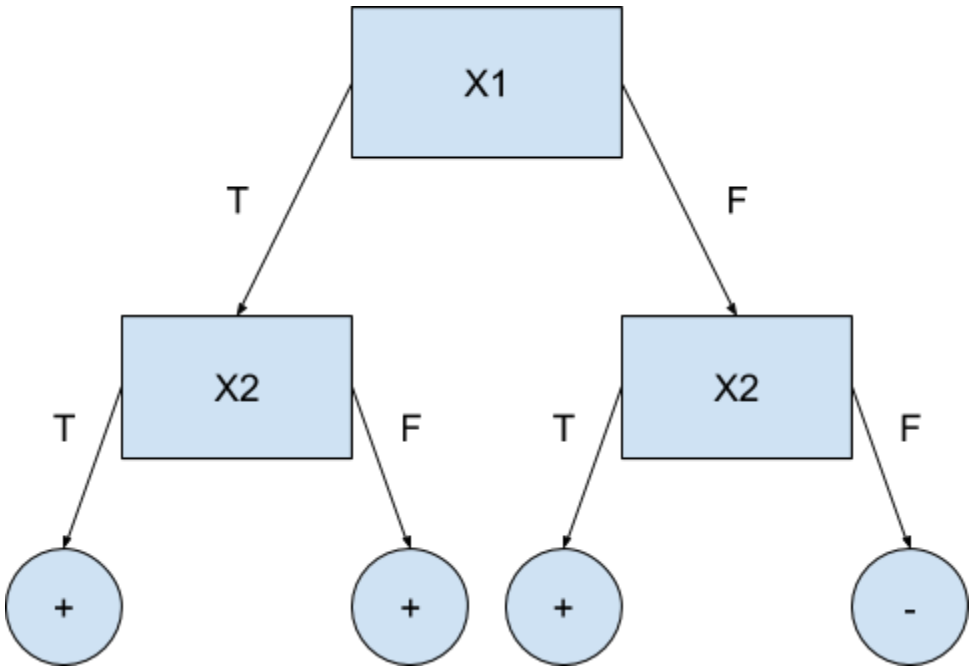
   Entropy of T → -(7/10)log(7/10)-(3/10)log(3/10) → .881
   Entropy of F → -(5/11)log(5/11)-(6/11)log(6/11) → .994

   IG(x2) = H(Y) - H(Y|x2) = .98522 - (10/21)(.881) - (11/21)(.994) = **.045**

   Variable x1 contains the larger information gain

3. ID3 tree

   The information gained decides which feature comes first.
   The final nodes are determined by ratio.
   TT is 100% +
   TF is 4:1 +
   FT is 4:3 +
   FF is 5:1 -


X1
T
F
X2
X2
T
F
T
F
+
+
+
-