Final Exam :

(Q1) Level 1 of impurity :

$I_p = P(E)*(1-P(E)) + (PM)*(1-P(M)) + P(T)*(1-P(T)) + P(G)*(1-P(G))$

$\Rightarrow (2/7)(1-2/7) + (1/7)(1-1/7) + (2/7)(1-2/7) + (2/7)(1-2/7)$

$\Rightarrow (0.28)(0.71) + (0.14)(0.85) + (0.28)(0.71) + (0.25)(0.71)$

$I_p \Rightarrow 0.7154$

Average $I_p \Rightarrow (1/7) * I_p$

$\Rightarrow 1 (0.7154)$

Level 2 : Impurity

Left subtree :

$I_p \Rightarrow P(T)*(1-P(T)) + P(G)*(1-P(G))$

$\Rightarrow (2/4)(1-2/4) + (2/4)(1-2/4)$

$I_{p_1} \Rightarrow 1/2 \Rightarrow 0.5$

Subtree 2:

$I_p \Rightarrow P(E)(1-P(E)) + P(M)(1-P(M))$

$\Rightarrow (2/3)(1-2/3) + (1/3)(1-1/3)$

$I_{p_2} \Rightarrow \approx 0.44$

Average $I_p \Rightarrow (4/7) I_{p_1} + (3/7) I_{p_2}$

$Avg\_I_p \Rightarrow (4/7)(0.5) + (3/7)(0.44)$

$\Rightarrow 0.4742$

Information gain in 2 levels:

info-gain => Avg I_p (Level 1) - Avg I_p (Level 2)

=> 0.7154 - 0.4743

=> 0.2411