(4) $H(\text{Output}) = -(P_0 \log_2 P_0 + P_1 \log_2 P_1) = -(-0.441 + -0.53)$

$= 0.97$ bits

$\text{Rem}(A_1) = \frac{1}{5}(0 + 1\log_2 1) + \frac{4}{5}(1) = 4/5 = \underline{0.8}$

$\text{Rem}(A_2) = \frac{2}{5}(0) + \frac{3}{5}\left(\frac{1}{3}\log_2 \frac{1}{3} + \frac{2}{3}\log_2 \frac{2}{3}\right) = \underline{0.551}$

$\text{Rem}(A_3) = \frac{3}{5}\left(\frac{2}{3}\log_2 \frac{2}{3} + \frac{1}{3}\log_2 \frac{1}{3}\right) + \frac{2}{5}(1) = 0.951$

Splitting on A_2 maximizes information gain