Submission for Apoorv Walia and Marco Lagos

A3 3.2

Mis-classification A = $\frac{100+100}{400+400} = \frac{1}{4}$

Misclassification B = $\frac{200}{400+400} = \frac{1}{4}$

$H(D) = -\frac{1}{2}\log(1/2) - \frac{1}{2}\log(1/2) = 1$

Entropy for A

$H(D_1) = -\frac{1}{4}\log(1/4) - \frac{1}{4}\log(1/4) = 0.811$

$H(D_2) = -\frac{1}{4}\log(1/4) - \frac{1}{4}\log(1/4) = 0.811$

Entropy gain = $1 - \frac{1}{2}(0.811) - \frac{1}{2}(0.811) = 0.189$

Entropy for B

$H(D_1) = -\frac{1}{3}\log(1/3) - \frac{2}{3}\log(2/3) = 0.918$

$H(D_2) = -\log(-1) = 0$

Entropy gain = $1 - \frac{3}{4}(0.918) = 0.3115$

Since Entropy gain B > entropy gain A, model B is the preferred split.

$G(D) = 2 \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{2}$

Gini Index for A

$G(D_1) = 2 \times \frac{1}{4} \times \frac{3}{4} = \frac{3}{8}$

$G(D_2) = 2 \times \frac{1}{4} \times \frac{3}{4} = \frac{3}{8}$

Gini Index = $\frac{1}{2} - \frac{1}{2} \times \frac{3}{8} + \frac{1}{2} \times \frac{3}{8} = \frac{1}{8}$

Gini Index for B

$$G(D_1) = 2 \times \frac{1}{3} \times \frac{2}{3} = \frac{4}{9}$$

$$G(D_2) = 0$$

$$\text{Gini Index} = \frac{1}{2} - \frac{3}{4} \times \frac{4}{9} = \frac{1}{6}$$

Since Gini Index B > Gini Index A, model B is the preferred split.