Homework - 3.

(1)

Entropy - is the measure of impurity in bunch of data.

Formula for Entropy = $\sum_i -p_i \log_2(p_i)$.

where i is the class.

In the Marital data from census, we have 3 classes from the one target which is (<25K, 25K - 50K, >50K).

Entropy =

$$-\sum_{i\{<25K,\ 25-50K,\ >50K\}} p_i \log_2(p_i)$$

$P_{<25} = \frac{2}{8}$

$P_{25K-50K} = \frac{5}{8}$

$P_{>50K} = \frac{1}{8}$

$$= -\left(\left(\frac{2}{8} \times \log_2\left(\frac{2}{8}\right)\right) + \left(\frac{5}{8} \times \log_2\left(\frac{5}{8}\right)\right) + \left(\frac{1}{8} \times \log_2\left(\frac{1}{8}\right)\right)\right)$$

$$= -\left(\left(\frac{2}{8} \times -2\right) + \left(\frac{5}{8} \times -0.678\right) + \left(\frac{1}{8} \times -3\right)\right)$$

$$= -\left(-0.5 - 0.423 - 0.375\right)$$

$$= -(-1.298) = \boxed{1.298 \text{ bits}}$$