3.

senior: 30 + 5 + 3 + 10 + 4 = 52

junior: 40 + 40 + 20 + 3 + 4 + 6 = 113

Total = 113 + 52 = 165

$$H(s) = -\left[\frac{52}{165} \cdot \log_2\left(\frac{52}{165}\right) + \frac{113}{165} \cdot \log\left(\frac{113}{165}\right)\right] = 0.899$$

Age "31-35" group is the only group contributing to the weighted entropy for age

$$H(age) = \frac{79}{165} \times \left[-\frac{44}{79} * \log_2\left(\frac{44}{79}\right) - \frac{35}{79} \times \log_2\left(\frac{35}{79}\right)\right]$$

$$= 0.4743$$

$$Gain(age) = H(S) - H(age) = 0.899 - 0.474 = 0.425$$

Department:

E_1sales: $-\frac{80}{110} \log_2\left(\frac{80}{110}\right) - \frac{30}{110} \log_2\left(\frac{30}{110}\right) = 0.8453$

E_2systems: $-\frac{23}{31} \log_2\left(\frac{23}{31}\right) - \frac{8}{31} \log_2\left(\frac{8}{31}\right) = 0.8238$

E_3marketing: $-\frac{4}{14} \log_2\left(\frac{4}{14}\right) - \frac{10}{14} \log_2\left(\frac{10}{14}\right) = 0.8631$

E_4secretary: $-\frac{6}{10} \log_2\left(\frac{6}{10}\right) - \frac{4}{10} \log_2\left(\frac{4}{10}\right) = 0.9710$

$$Gain(salary) = H(S) - \left(\frac{110}{165} \times E_1 + \frac{31}{165} \times E_2 + \frac{14}{165} \times E_3 + \frac{10}{165} \times E_4\right)$$

$$= 0.899 - (0.5636 + 0.1548 + 0.0732 + 0.0588)$$

$$\approx 0.049$$

should be the initial split point in the decision tree.

$$Gain(salary) = H(S) - \frac{63}{165} \times \left[-\frac{23}{63} \log_2\left(\frac{23}{63}\right) - \frac{40}{63} \log_2\left(\frac{40}{63}\right)\right]$$

$$\approx 0.538$$

So, the salary attribute has the highest information gain.