Information gain

- attribute selection measure
- Claude Shannon on information theory

* The expected information needed to classify a tuple in D is given by

$$Info(D) = -\sum_{i=1}^{m} p_i \log_2(p_i)$$

p_i is the probability that an arbitrary tuple in D belongs to class C_i and is estimated by $|C_{i,D}| / |D|$

info(D) is also known as the entropy of D

$$Info_A(D) = \sum_{j=1}^{v} \frac{|D_j|}{|D|} \times Info(D_j)$$

Information gain is defined as the difference between the original information requirement and the new requirement after partitioning on A attribute

$$Gain(A) = Info(D) - Info_A(D)$$

$$Info(D) = -\frac{9}{14}\log_2\left(\frac{9}{14}\right) - \frac{5}{14}\log_2\left(\frac{5}{14}\right)$$

$$= 0.940$$

$$Info_{age}(D) = \frac{5}{14} \times \left(-\frac{2}{5}\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\log\left(\frac{3}{5}\right)\right)$$
$$+ \frac{4}{14} \times \left(-\frac{4}{4}\log_2\left(\frac{4}{4}\right) - \frac{0}{4}\log\left(\frac{0}{4}\right)\right)$$
$$+ \frac{5}{14}\left(-\frac{3}{5}\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\log_2\left(\frac{2}{5}\right)\right)$$

$$= 0.694$$

$$Gain(age) = Info(D) - Info_{age}(D) = 0.940 - 0.694 = 0.246 \text{ bits}$$

Gain(income) = 0.029 bits ; Gain(student) = 0.151 b

Gain(credit_rating) = 0.048 bits.

Age has the highest information gain among the attributes selected for splitting.