

2013

1) A medida usada pelo ID3 para a escolha do melhor atributo é a entropia, enquanto que o C4.5 usa a razão do ganho.

O C4.5 é capaz de lidar com missing values e valores de atributos contínuos, o que não é verdade para o ID3.

O C4.5 também pode a árvore, eliminando sub-árvores que não contribuem o suficiente para o ganho da informação, o que não é verdade para o ID3.

2)

2012

1) Igual a 2013

2) m22(60)
22e25(30)
 π 25(10)

$$\text{info}(\text{Cetária}) = -\frac{60}{100} \times \log_2\left(\frac{60}{100}\right)$$

$$- \frac{30}{100} \times \log_2\left(\frac{30}{100}\right)$$

$$- \frac{10}{100} \times \log_2\left(\frac{10}{100}\right) = 1.14 \text{ bits}$$

3) e1d(80) \rightarrow m22(60)
fc(20) 22e25(20)
 π 25(0)
 \downarrow
22e25(10)
 π 25(10)
m22(0)

$$\text{info}(C/\text{idade}) = \frac{80}{100} \times \left[-\frac{60}{80} \times \log_2\left(\frac{60}{80}\right) - \frac{10}{80} \times \log_2\left(\frac{10}{80}\right) \right]$$

$$+ \frac{20}{100} \times \left[-\frac{10}{20} \times \log_2\left(\frac{10}{20}\right) - \frac{10}{20} \times \log_2\left(\frac{10}{20}\right) \right]$$

$$= 0.749 \text{ bits}$$