

$$Entropy = \sum_{i=1}^c -p_i \log_2 p_i$$

$p = \frac{n. \text{ de vezes que o valor aparece}}{\text{Total de registos}}$

$i = \text{cada valor que o atributo pode assumir}$

$$Gain(S, A) = Entropy(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

Risco
Alto
Alto
Moderado
Alto
Baixo
Baixo
Alto
Moderado
Baixo
Baixo
Alto
Moderado
Baixo
Alto

Alto: 6/14

Moderado: 3/14

Baixo: 5/14

→ Calcular Entropia

$$E(S) = -\frac{6}{14} \cdot \log_2 \frac{6}{14}$$

$$-\frac{3}{14} \cdot \log_2 \frac{3}{14}$$

$$-\frac{5}{14} \cdot \log_2 \frac{5}{14}$$

$$= 1,53$$

Entropia do Classificador

→ Ganho ao histórico de crédito

História do crédito	Risco
Ruim	Alto
Desconhecida	Alto
Desconhecida	Moderado
Desconhecida	Alto
Desconhecida	Baixo
Desconhecida	Baixo
Ruim	Alto
Ruim	Moderado
Boa	Baixo
Boa	Baixo
Boa	Alto
Boa	Moderado
Boa	Baixo
Ruim	Alto

Bom / Alto 1/5
5/14 - Moderado 1/5
Ruim 3/5

$$E(S) = -1/5 \cdot \log_2 \frac{1}{5} - \frac{1}{5} \log_2 \frac{1}{5} - \frac{3}{5} \cdot \log_2 \frac{3}{5} = 1,37$$

14 - Desconhecida / Alto 2/5
5/14 - Moderado 1/5
Ruim 4/14 - Baixo 2/5
Alto 3/4
Moderado 1/4
Baixo 0

$$E(S) = 1,52$$

História de crédito

→ Calcular a entropia Bom, Desconhecida e ruim

→ Aplicar a fórmula do ganho

$$G(\text{história}) = 1,53 - \left(\frac{5}{14} \cdot 1,37 \right) - \left(\frac{5}{14} \cdot 1,52 \right) - \left(\frac{4}{14} \cdot 0,81 \right) = 0,26$$