

1. ENTROPIA TOTAL

S = 17 exemplos

Classe	Quantidade
Bola	5 exemplos
Losango	6 exemplos
Quadrado	6 Exemplos

$$E(s) = \left(-\frac{5}{17} \log_2 \frac{5}{17}\right) + \left(-\frac{6}{17} \log_2 \frac{6}{17}\right) + \left(-\frac{6}{17} \log_2 \frac{6}{17}\right)$$

$$E(s) = 0.519 + 0.530 + 0.530 = 1.579$$

2. GANHOS

Para X1:

Figura 1 - Valores e classes para atributo X1

x1	classificacao
0	bola
0	quadrado
1	bola
2	bola
2	quadrado
4	bola
4	quadrado
4	quadrado
4	quadrado
6	losango
8	bola
8	losango
8	losango
8	quadrado
9	losango
9	losango
10	losango

fonte: Próprio autor (2022)

Atributo para particionar = valor ≤ 5

$V = \{A = \text{valores} \leq 5; B = \text{valores} > 5\}$

$S_V - S_{VA} = 9$ exemplos, sendo 4 bolas e 5 quadrados;

$S_{VB} = 8$ exemplos, sendo 6 losangos, 1 quadrado e 1 bola;

$$E(S_{VA}) = \left(-\frac{4}{9} \log_2 \frac{4}{9}\right) + \left(-\frac{5}{9} \log_2 \frac{5}{9}\right) = 0.519 + 0.471 = 0.990$$

$$E(S_{VB}) = \left(-\frac{1}{8} \log_2 \frac{1}{8}\right) + \left(-\frac{6}{8} \log_2 \frac{6}{8}\right) + \left(-\frac{1}{8} \log_2 \frac{1}{8}\right) = 0.375 + 0.311 + 0.375 = 1.061$$

Portanto,

$$G(S, X1) = 1.579 - \left(\frac{9}{17} * 0.990 + \frac{8}{17} * 1.061\right) = 0.556$$

Para X2:

Figura 2 - Valores e classes para atributo X2

x2	classificacao
1	bola
1	bola
3	bola
3	losango
3	losango
4	bola
4	losango
5	bola
5	losango
5	losango
7	quadrado
9	quadrado
9	quadrado
9	losango
10	quadrado
10	quadrado
10	quadrado

fonte: Próprio autor (2022)

Atributo para particionar = valor ≤ 6

$V = \{A = \text{valores} \leq 6; B = \text{valores} > 6\}$

$S_V - S_{VA} = 10$ exemplos, sendo 5 bolas e 5 losangos;

$S_{vB} = 7$ exemplos, sendo 6 quadrados e 1 losango;

$$E(S_{vA}) = \left(-\frac{5}{10} \log_2 \frac{5}{10}\right) + \left(-\frac{5}{10} \log_2 \frac{5}{10}\right) = 0.5 + 0.5 = 1$$

$$E(S_{vB}) = \left(-\frac{6}{7} \log_2 \frac{6}{7}\right) + \left(-\frac{1}{7} \log_2 \frac{1}{7}\right) = 0.190 + 0.401 = 0.591$$

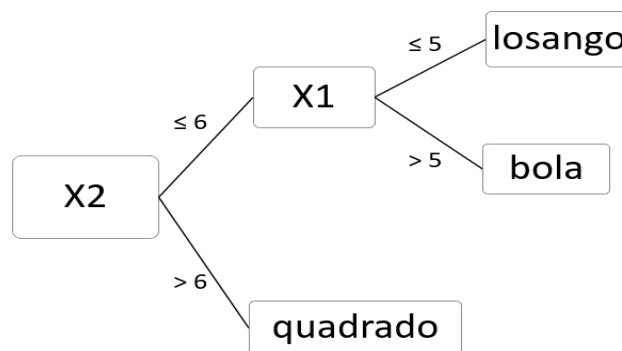
Portanto,

$$G(S, X2) = 1.579 - \left(\frac{10}{17} * 1 + \frac{7}{17} * 0.591\right) = 0.747$$

Sendo assim, o atributo de maior ganho é o x2

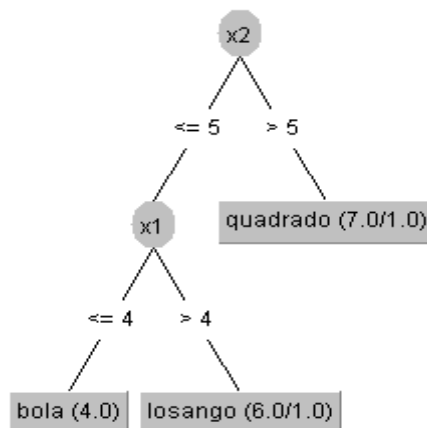
3. ÁRVORES DE DECISÃO

Figura 3 - Árvore de decisão elaborada no passo-a-passo



fonte: Próprio autor (2022)

Figura 4 - Árvore de decisão elaborada pelo programa Weka utilizando o algoritmo J48



fonte: Weka

4. RESULTADOS E DISCUSSÃO

Foi utilizado o algoritmo J48 por trabalhar com atributos com valores numéricos e assim determinar limites para classificação. Para a classificação manual foram determinados os valores de ocorrência para cada classe e escolhido um limiar. Para os valores de X2, por exemplo, há ocorrência de bola e losango para 1, 3, 4 e 5, e há ocorrência de quadrado e losango para 7, 9 e 10, ao desconsiderar o losango, o limiar está entre 5 e 7, portanto 6 ($(5+7)/2$).

Quadro 1 - comparativo de resultados

Resultado esperado	J48 - Weka		Manual	
bola	bola	ok	bola	ok
bola	bola	ok	bola	ok
losango	losango	ok	losango	ok
losango	quadrado	x	losango	ok
quadrado	quadrado	ok	quadrado	ok
quadrado	quadrado	ok	quadrado	ok
bola	bola	ok	bola	ok
bola	losango	x	losango	x
bola	quadrado	x	bola	ok
quadrado	quadrado	ok	quadrado	ok
losango	quadrado	x	losango	ok
bola	losango	x	bola	ok
bola	losango	x	losango	x
bola	quadrado	x	bola	ok
quadrado	quadrado	ok	quadrado	ok
quadrado	quadrado	ok	bola	x
losango	losango	ok	bola	x
losango	losango	ok	losango	ok

fonte: Próprio autor (2022)

Para os valores verificados, o programa retornou 65% de acerto e a árvore manual obteve 82% de acerto. É necessário considerar que há pontos que não tem o comportamento esperado, e, portanto, incrementam a desordem do grupo amostral. O ideal seria retirar os pontos para que houvesse uma melhor classificação da árvore de decisão.