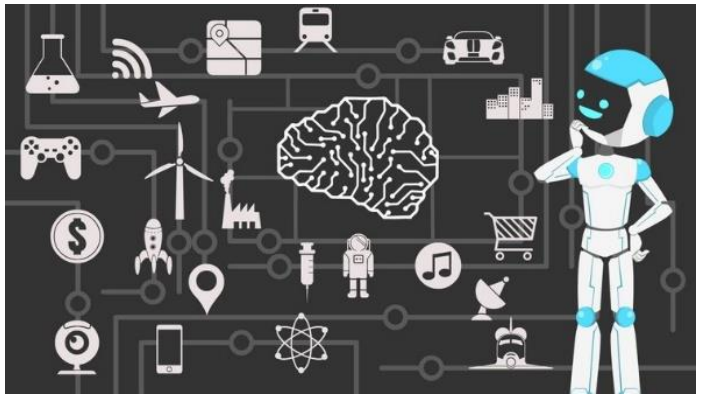


FORMAÇÃO INTELIGÊNCIA ARTIFICIAL E MACHINE LEARNING

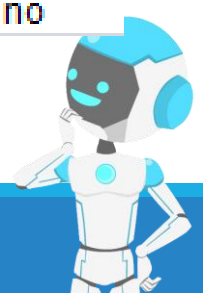
MACHINE LEARNING – ESTUDANDO ALGORITMOS

ÁRVOES DE DECISÃO PARTE II

Criando a Árvore de Decisão?

No.	1: outlook	2: temperature	3: humidity	4: windy	5: **play**
	Nominal	Nominal	Nominal	Nominal	Nominal
1	sunny	hot	high	FALSE	no
2	sunny	hot	high	TRUE	no
3	overcast	hot	high	FALSE	yes
4	rainy	mild	high	FALSE	yes
5	rainy	cool	normal	FALSE	yes
6	rainy	cool	normal	TRUE	no
7	overcast	cool	normal	TRUE	yes
8	sunny	mild	high	FALSE	no
9	sunny	cool	normal	FALSE	yes
10	rainy	mild	normal	FALSE	yes
11	sunny	mild	normal	TRUE	yes
12	overcast	mild	high	TRUE	yes
13	overcast	hot	normal	FALSE	yes
14	rainy	mild	high	TRUE	no

- Temos 4 atributos “candidatos” a ser o nó raiz.
- Qual devemos escolher?
  - Buscar o que tenha maior ganho de informação!



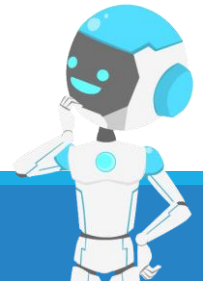
Entropia

Teoria da Informação

$$E(S) = -\sum_{i=1}^{n} p_i \log_2 pi$$

Se todas as instancias de S pertencem a mesma classe E(S) = 0

Se S contem o mesmo número de instancia para cada classe, E(s) = 1



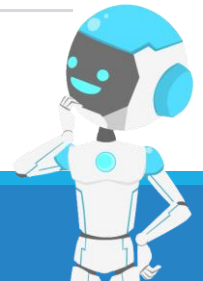
Cálculo da Entropia - Classe

$$E(S) = -\sum_{i=1}^{n} p_i \log_2 pi$$

$$E(S) = \left(-\frac{9}{14}\log_2\left(\frac{9}{14}\right)\right) + \left(-\frac{5}{14}\log_2\left(\frac{5}{14}\right)\right)$$

$$E(S) = 0{,}94$$

No.	1: outlook	2: temperature	3: humidity	4: windy	5: **play**
	Nominal	Nominal	Nominal	Nominal	Nominal
1	sunny	hot	high	FALSE	no
2	sunny	hot	high	TRUE	no
3	overcast	hot	high	FALSE	yes
4	rainy	mild	high	FALSE	yes
5	rainy	cool	normal	FALSE	yes
6	rainy	cool	normal	TRUE	no
7	overcast	cool	normal	TRUE	yes
8	sunny	mild	high	FALSE	no
9	sunny	cool	normal	FALSE	yes
10	rainy	mild	normal	FALSE	yes
11	sunny	mild	normal	TRUE	yes
12	overcast	mild	high	TRUE	yes
13	overcast	hot	normal	FALSE	yes
14	rainy	mild	high	TRUE	no



Cálculo da Entropia - Outlook

$$E(S) = -\sum_{i=1}^{n} p_i \log_2 pi$$

Outlook (sunny – para yes e no)

$$E(S) = \left(-\frac{2}{5}\log_2\left(\frac{2}{5}\right)\right) + \left(-\frac{3}{5}\log_2\left(\frac{3}{5}\right)\right) = 0{,}97$$

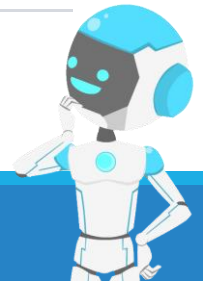
Outlook (overcast – para yes e no)

$$E(S) = \left(-\frac{4}{4}\log_2\left(\frac{4}{4}\right)\right) + \left(-\frac{0}{4}\log_2\left(\frac{0}{4}\right)\right) = 0$$

Outlook (rainy – para yes e no)

$$E(S) = \left(-\frac{3}{5}\log_2\left(\frac{3}{5}\right)\right) + \left(-\frac{2}{5}\log_2\left(\frac{2}{5}\right)\right) = 0{,}97$$

No.	1: outlook	2: temperature	3: humidity	4: windy	5: **play**
	Nominal	Nominal	Nominal	Nominal	Nominal
1	sunny	hot	high	FALSE	no
2	sunny	hot	high	TRUE	no
3	overcast	hot	high	FALSE	yes
4	rainy	mild	high	FALSE	yes
5	rainy	cool	normal	FALSE	yes
6	rainy	cool	normal	TRUE	no
7	overcast	cool	normal	TRUE	yes
8	sunny	mild	high	FALSE	no
9	sunny	cool	normal	FALSE	yes
10	rainy	mild	normal	FALSE	yes
11	sunny	mild	normal	TRUE	yes
12	overcast	mild	high	TRUE	yes
13	overcast	hot	normal	FALSE	yes
14	rainy	mild	high	TRUE	no



Cálculo da Entropia - Temperature

$$E(S) = -\sum_{i=1}^{n} p_i \log_2 pi$$

Temperature (hot – para yes e no)

$$E(S) = \left(-\frac{2}{4}\log_2\left(\frac{2}{4}\right)\right) + \left(-\frac{2}{4}\log_2\left(\frac{2}{4}\right)\right) = 1$$

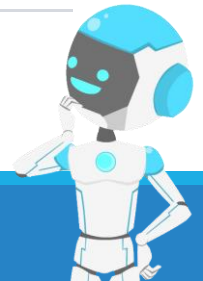
Temperature (mild – para yes e no)

$$E(S) = \left(-\frac{4}{6}\log_2\left(\frac{4}{6}\right)\right) + \left(-\frac{2}{6}\log_2\left(\frac{2}{6}\right)\right) = 0{,}91$$

Temperature (cold – para yes e no)

$$E(S) = \left(-\frac{3}{4}\log_2\left(\frac{3}{4}\right)\right) + \left(-\frac{1}{4}\log_2\left(\frac{1}{4}\right)\right) = 0{,}81$$

No.	1: outlook	2: temperature	3: humidity	4: windy	5: **play**
	Nominal	Nominal	Nominal	Nominal	Nominal
1	sunny	hot	high	FALSE	no
2	sunny	hot	high	TRUE	no
3	overcast	hot	high	FALSE	yes
4	rainy	mild	high	FALSE	yes
5	rainy	cool	normal	FALSE	yes
6	rainy	cool	normal	TRUE	no
7	overcast	cool	normal	TRUE	yes
8	sunny	mild	high	FALSE	no
9	sunny	cool	normal	FALSE	yes
10	rainy	mild	normal	FALSE	yes
11	sunny	mild	normal	TRUE	yes
12	overcast	mild	high	TRUE	yes
13	overcast	hot	normal	FALSE	yes
14	rainy	mild	high	TRUE	no



Cálculo da Entropia - Humidity

$$E(S) = -\sum_{i=1}^{n} p_i \log_2 pi$$

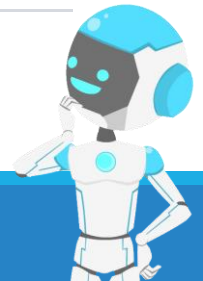
Humidity (high – para yes e no)

$$E(S) = \left(-\frac{3}{7}\log_2\left(\frac{3}{7}\right)\right) + \left(-\frac{4}{7}\log_2\left(\frac{4}{7}\right)\right) = 0{,}98$$

Humidity (normal – para yes e no)

$$E(S) = \left(-\frac{6}{7}\log_2\left(\frac{6}{7}\right)\right) + \left(-\frac{1}{7}\log_2\left(\frac{1}{7}\right)\right) = 0{,}59$$

No.	1: outlook	2: temperature	3: humidity	4: windy	5: **play**
	Nominal	Nominal	Nominal	Nominal	Nominal
1	sunny	hot	high	FALSE	no
2	sunny	hot	high	TRUE	no
3	overcast	hot	high	FALSE	yes
4	rainy	mild	high	FALSE	yes
5	rainy	cool	normal	FALSE	yes
6	rainy	cool	normal	TRUE	no
7	overcast	cool	normal	TRUE	yes
8	sunny	mild	high	FALSE	no
9	sunny	cool	normal	FALSE	yes
10	rainy	mild	normal	FALSE	yes
11	sunny	mild	normal	TRUE	yes
12	overcast	mild	high	TRUE	yes
13	overcast	hot	normal	FALSE	yes
14	rainy	mild	high	TRUE	no



Cálculo da Entropia - Windy

$$E(S) = -\sum_{i=1}^{n} p_i \log_2 pi$$

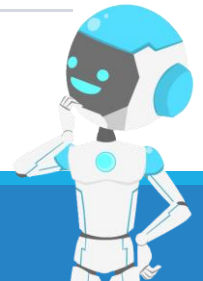
Windy (True – para yes e no)

$$E(S) = \left(-\frac{3}{6}\log_2\left(\frac{3}{6}\right)\right) + \left(-\frac{3}{6}\log_2\left(\frac{3}{6}\right)\right) = 1$$

Windy (False – para yes e no)

$$E(S) = \left(-\frac{6}{8}\log_2\left(\frac{6}{8}\right)\right) + \left(-\frac{2}{8}\log_2\left(\frac{2}{8}\right)\right) = 0{,}81$$

No.	1: outlook	2: temperature	3: humidity	4: windy	5: **play**
	Nominal	Nominal	Nominal	Nominal	Nominal
1	sunny	hot	high	FALSE	no
2	sunny	hot	high	TRUE	no
3	overcast	hot	high	FALSE	yes
4	rainy	mild	high	FALSE	yes
5	rainy	cool	normal	FALSE	yes
6	rainy	cool	normal	TRUE	no
7	overcast	cool	normal	TRUE	yes
8	sunny	mild	high	FALSE	no
9	sunny	cool	normal	FALSE	yes
10	rainy	mild	normal	FALSE	yes
11	sunny	mild	normal	TRUE	yes
12	overcast	mild	high	TRUE	yes
13	overcast	hot	normal	FALSE	yes
14	rainy	mild	high	TRUE	no



Ganho de Informação (Information Gain)

$$IG(S, A) = E(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} E(S_i)$$

Outlook

sunny overcast rainy

$$IG(S, A) = 0{,}94 - \frac{5}{14} * 0{,}97 - \frac{4}{14} * 0 - \frac{5}{14} * 0{,}97 = 0{,}2471$$

Entropia Classe

Proporção sunny

Entropia sunny

Proporção overcast

Entropia overcast

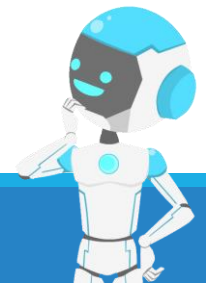
Proporção rainy

Entropia rainy

$IG(S, A) = E(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} E(S_i)$

Outlook

$IG(S, A) = 0{,}94 - \frac{5}{14} * 0{,}97 - \frac{4}{14} * 0 - \frac{5}{14} * 0{,}97 = 0{,}2471$



Ganho de Informação (Information Gain)

$$\text{IG}(S,A) = Entropia(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} Entropia(S_i)$$

Outlook

$$\text{IG}(S,A) = 0{,}94 \; -\frac{5}{14} * 097 \; -\frac{4}{14} * 0 \; -\frac{5}{14} * 0{,}97 = 0{,}2471$$

Temperatures

$$\text{IG}(S,A) = 0{,}94 \; -\frac{4}{14} * 1 \; -\frac{6}{14} * 0{,}91 \; -\frac{4}{14} * 0{,}81 = \; 0{,}0328$$

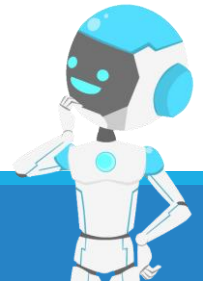
Humidy

$$\text{IG}(S,A) = 0{,}94 \; -\frac{7}{14} * 0{,}97 \; -\frac{7}{14} * 0{,}59 \; = 0{,}16$$

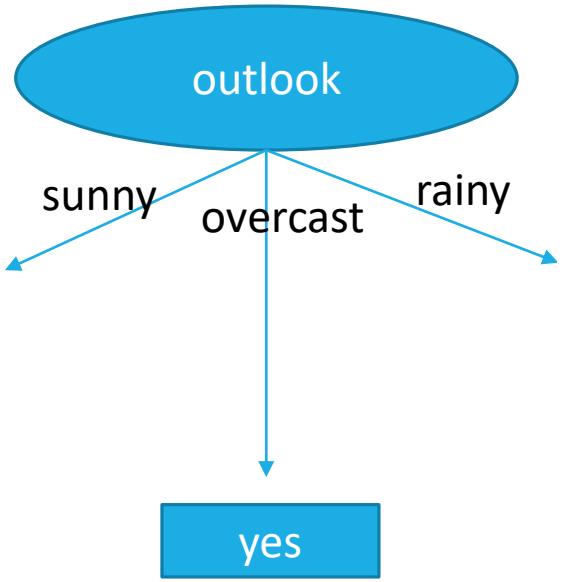
Windy

$$\text{IG}(S,A) = 0{,}94 \; -\frac{6}{14} * 1 \; -\frac{8}{14} * 0{,}81 \; = 0{,}048$$

No.	1: outlook	2: temperature	3: humidity	4: windy	5: **play**
	Nominal	Nominal	Nominal	Nominal	Nominal
1	sunny	hot	high	FALSE	no
2	sunny	hot	high	TRUE	no
3	overcast	hot	high	FALSE	yes
4	rainy	mild	high	FALSE	yes
5	rainy	cool	normal	FALSE	yes
6	rainy	cool	normal	TRUE	no
7	overcast	cool	normal	TRUE	yes
8	sunny	mild	high	FALSE	no
9	sunny	cool	normal	FALSE	yes
10	rainy	mild	normal	FALSE	yes
11	sunny	mild	normal	TRUE	yes
12	overcast	mild	high	TRUE	yes
13	overcast	hot	normal	FALSE	yes
14	rainy	mild	high	TRUE	no



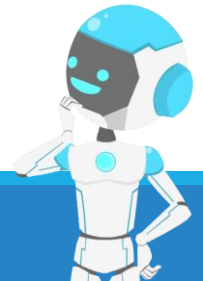
Primeiro nodo



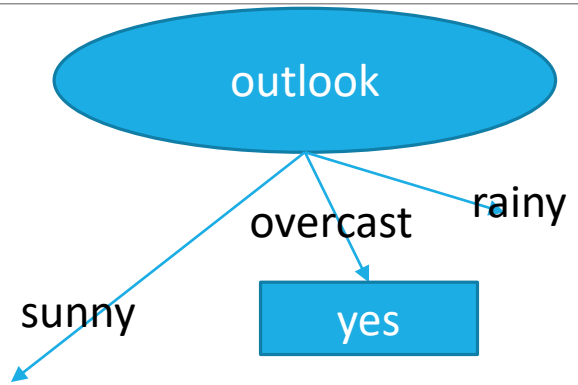
outlook	temperature	humidity	windy	play
sunny	hot	high	FALSE	no
sunny	hot	high	TRUE	no
sunny	mild	high	FALSE	no
sunny	cool	normal	FALSE	yes
sunny	mild	normal	TRUE	yes

outlook	temperature	humidity	windy	play
rainy	mild	high	FALSE	yes
rainy	cool	normal	FALSE	yes
rainy	cool	normal	TRUE	no
rainy	mild	normal	FALSE	yes
rainy	mild	high	TRUE	no

outlook	temperature	humidity	windy	play
overcast	hot	high	FALSE	yes
overcast	cool	normal	TRUE	yes
overcast	mild	high	TRUE	yes
overcast	hot	normal	FALSE	yes



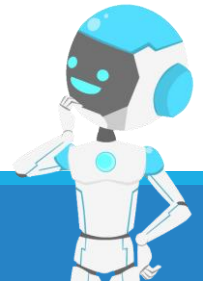
Particionando sunny



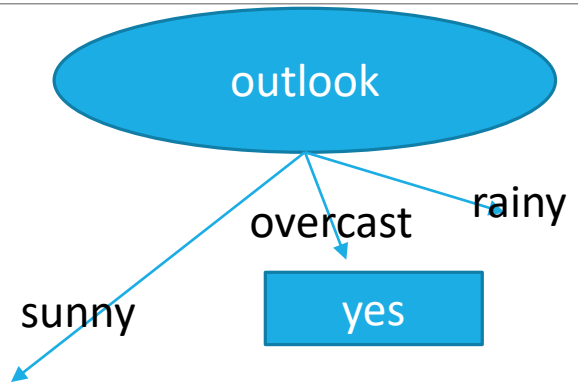
outlook	temperature	humidity	windy	play
sunny	hot	high	FALSE	no
sunny	hot	high	TRUE	no
sunny	mild	high	FALSE	no
sunny	cool	normal	FALSE	yes
sunny	mild	normal	TRUE	yes

Entropia Classe

$$E(sunny) = \left(-\frac{3}{5}\log_2\left(\frac{3}{5}\right)\right) + \left(-\frac{2}{5}\log_2\left(\frac{2}{5}\right)\right) = 0{,}97$$



Cálculo da Entropia - Temperature



outlook	temperature	humidity	windy	play
sunny	hot	high	FALSE	no
sunny	hot	high	TRUE	no
sunny	mild	high	FALSE	no
sunny	cool	normal	FALSE	yes
sunny	mild	normal	TRUE	yes

Temperatura (hot para yes e no)

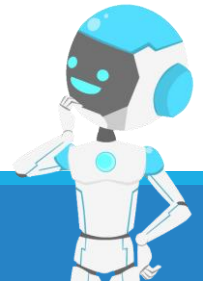
$$E(S) = \left(-\frac{0}{2}\log_2\left(\frac{0}{2}\right)\right) + \left(-\frac{2}{2}\log_2\left(\frac{2}{2}\right)\right) = 0$$

Temperatura (mild para yes e no)

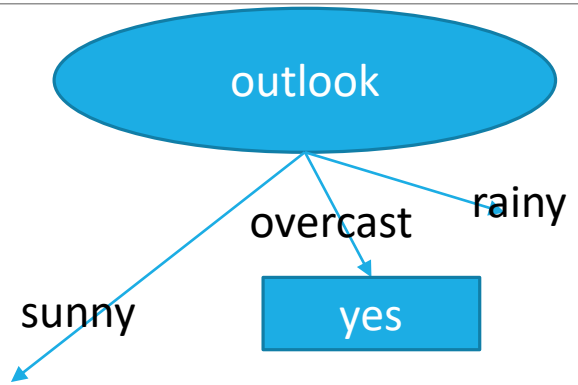
$$E(S) = \left(-\frac{1}{2}\log_2\left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2}\log_2\left(\frac{1}{2}\right)\right) = 1$$

Temperatura (cool para yes e no)

$$E(S) = \left(-\frac{1}{1}\log_2\left(\frac{1}{1}\right)\right) + \left(-\frac{0}{1}\log_2\left(\frac{0}{1}\right)\right) = 0$$



Cálculo da Entropia - Humidity



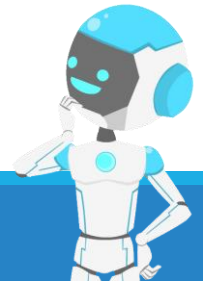
outlook	temperature	humidity	windy	play
sunny	hot	high	FALSE	no
sunny	hot	high	TRUE	no
sunny	mild	high	FALSE	no
sunny	cool	normal	FALSE	yes
sunny	mild	normal	TRUE	yes

Humidity (high para yes e no)

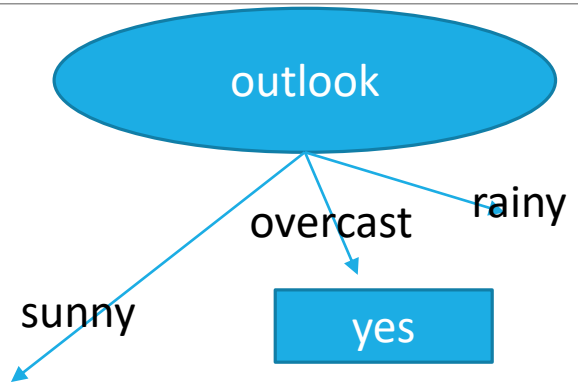
$$E(S) = \left(-\frac{0}{3}\log_2\left(\frac{0}{3}\right)\right) + \left(-\frac{3}{3}\log_2\left(\frac{3}{3}\right)\right) = 0$$

Humidity (normal para yes e no)

$$E(S) = \left(-\frac{2}{2}\log_2\left(\frac{2}{2}\right)\right) + \left(-\frac{0}{2}\log_2\left(\frac{0}{2}\right)\right) = 0$$



Cálculo da Entropia - Windy



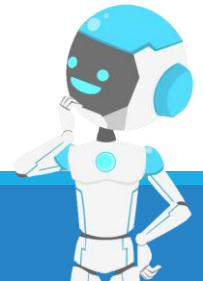
outlook	temperature	humidity	windy	play
sunny	hot	high	FALSE	no
sunny	hot	high	TRUE	no
sunny	mild	high	FALSE	no
sunny	cool	normal	FALSE	yes
sunny	mild	normal	TRUE	yes

windy (False para yes e no)

$$E(S) = \left(-\frac{1}{3}\log_2\left(\frac{1}{3}\right)\right) + \left(-\frac{2}{3}\log_2\left(\frac{2}{3}\right)\right) = 0{,}91$$

Humidity (True para yes e no)

$$E(S) = \left(-\frac{1}{2}\log_2\left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2}\log_2\left(\frac{1}{2}\right)\right) = 1$$



Ganho de Informação (Information Gain)

$$\text{IG}(S, A) = Entropia(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} Entropia(S_i)$$

Temperatures

$$\text{IG}(S, A) = 0{,}97 \; - \frac{2}{5} * 0 \; - \frac{2}{5} * 1 \; - \frac{1}{5} * 0 = 0{,}57$$

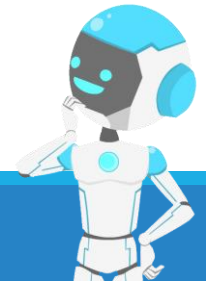
Humidity

$$\text{IG}(S, A) = 0{,}97 \; - \frac{3}{5} * 0 \; - \frac{2}{5} * 0 \; = 0{,}97$$

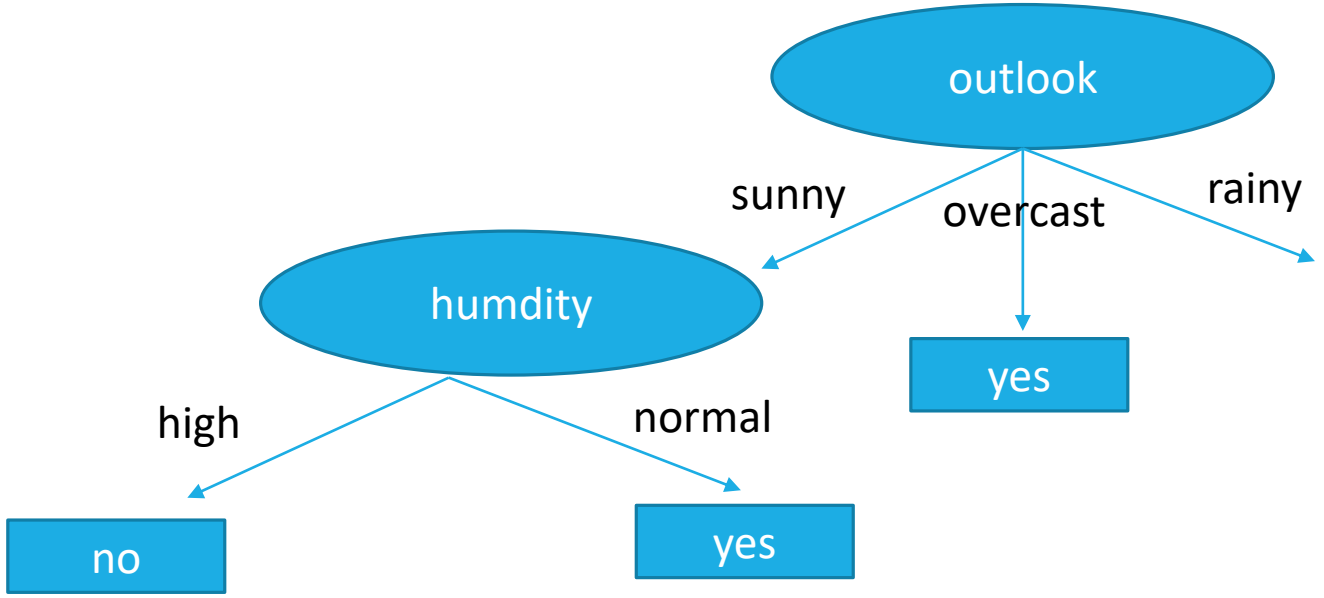
Windy

$$\text{IG}(S, A) = 0{,}97 \; - \frac{3}{5} * 0{,}91 \; - \frac{2}{5} * 1 \; = 0{,}024$$

outlook	temperature	humidity	windy	play
sunny	hot	high	FALSE	no
sunny	hot	high	TRUE	no
sunny	mild	high	FALSE	no
sunny	cool	normal	FALSE	yes
sunny	mild	normal	TRUE	yes

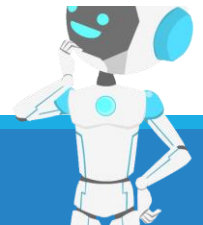


Próximo nodo

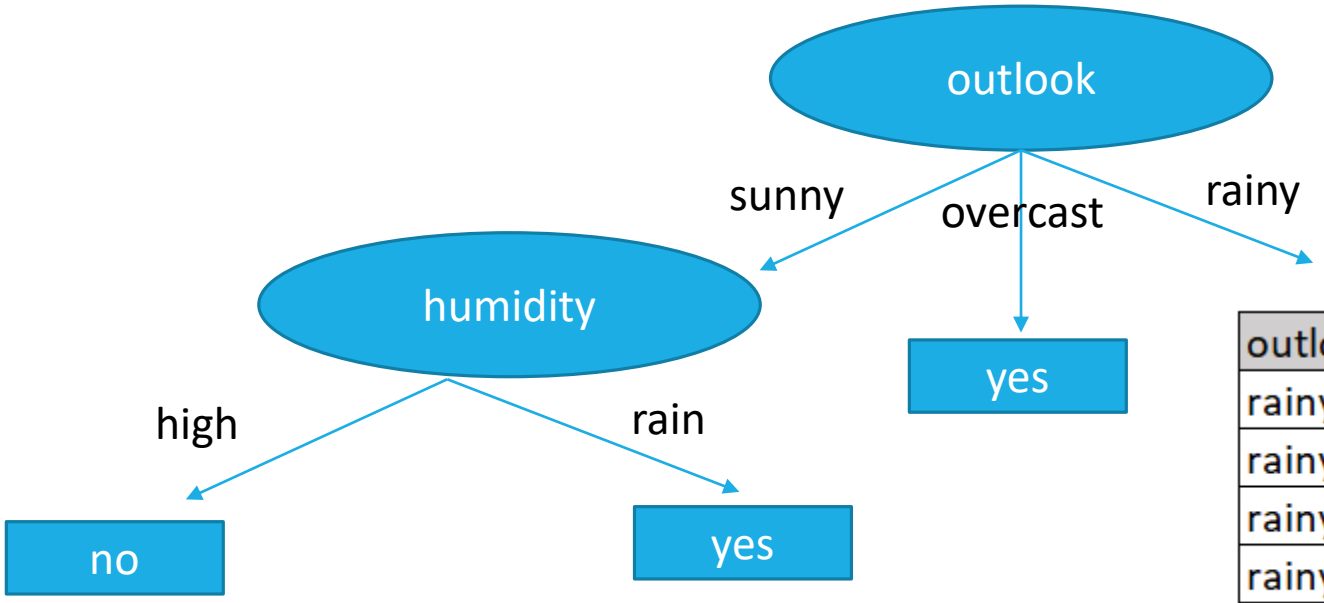


outlook	temperature	humidity	windy	play
sunny	hot	high	FALSE	no
sunny	hot	high	TRUE	no
sunny	mild	high	FALSE	no

outlook	temperature	humidity	windy	play
sunny	cool	normal	FALSE	yes
sunny	mild	normal	TRUE	yes

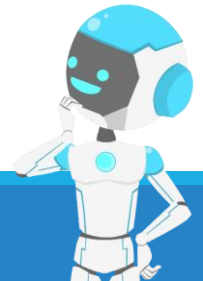


Continuando

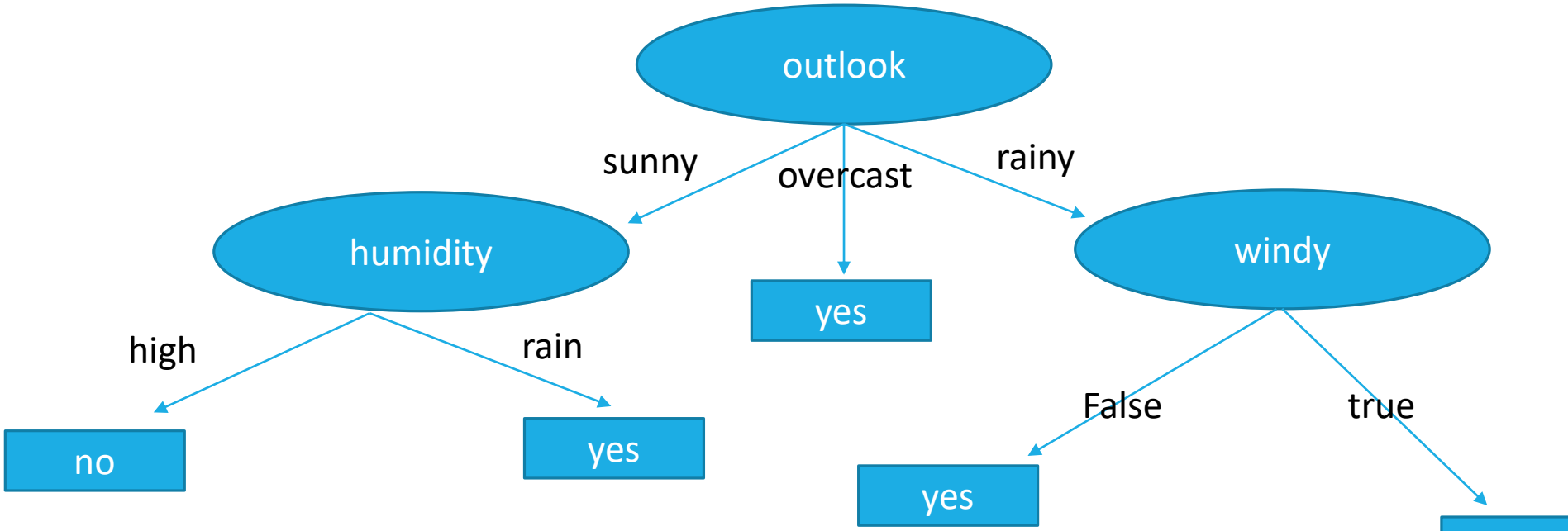


outlook	temperature	humidity	windy	play
rainy	mild	high	FALSE	yes
rainy	cool	normal	FALSE	yes
rainy	cool	normal	TRUE	no
rainy	mild	normal	FALSE	yes
rainy	mild	high	TRUE	no

Windy = 0,97



Continuando



outlook	temperature	humidity	windy	play
rainy	mild	high	FALSE	yes
rainy	cool	normal	FALSE	yes
rainy	mild	normal	FALSE	yes

outlook	temperature	humidity	windy	play
rainy	cool	normal	TRUE	no
rainy	mild	high	TRUE	no

Finalizando