Ministère de l'Enseignement supérieur et de la recherche scientifique
Université Abderrahmane Mira de Bejaïa
Faculté des Sciences Exactes
Département de Mathématiques

Module: Data Mining **Année: 2022-2023**
Nature de document: TD1-Solution
Niveau: L3-STID **N.BERMAD**

Exercice 3:

1. Déroulement de l'algorithme ID3 :

- **L'entropie de la population:**

$$E(C_{Jouer= Oui}, C_{Jouer= Non}) = -\left(\frac{9}{14}\log_2\left(\frac{9}{14}\right) + \frac{5}{14}\log_2\left(\frac{5}{14}\right)\right)$$
$$= -\left(0.64\ \frac{\ln\left(\frac{9}{14}\right)}{\ln(2)} + 0.35\ \frac{\ln\left(\frac{5}{14}\right)}{\ln(2)}\right) = 0.92$$

- **Calcul du gain de chaque attribut:**

A. Calcul de gain du Ciel:

1. Ciel= Ensoleillé:

$$\begin{cases} C_{Jouer= Oui} = \frac{2}{5} \\ C_{Jouer= Non} = \frac{3}{5} \end{cases} \qquad \frac{|\text{Ciel= Ensoleillé}|}{|\ C_{Jouer= Oui}, C_{Jouer= Non}\ |} = \frac{5}{14}$$

$$E(Ciel = Ensoleillé) = -\left(\frac{2}{5}\log_2\left(\frac{2}{5}\right) + \frac{3}{5}\log_2\left(\frac{3}{5}\right)\right) = 0.96$$ /* Entropie de l'attribut

2. Ciel= Pluie:

$$\begin{cases} C_{Jouer= Oui} = \frac{3}{5} \\ C_{Jouer= Non} = \frac{2}{5} \end{cases} \qquad \frac{|\text{Ciel= Pluie}|}{|\ C_{Jouer= Oui}, C_{Jouer= Non}\ |} = \frac{5}{14}$$

$$E(Ciel = Pluie) = -\left(\frac{3}{5}\log_2\left(\frac{3}{5}\right) + \frac{2}{5}\log_2\left(\frac{2}{5}\right)\right) = 0.96$$

3. Ciel=Couvert:

$$\begin{cases} C_{Jouer= Oui} = \frac{4}{4} = 1 \\ C_{Jouer= Non} = 0 \end{cases} \qquad \frac{|\text{Ciel= Couvert}|}{|\ C_{Jouer= Oui}, C_{Jouer= Non}\ |} = \frac{4}{14}$$

$$E(Ciel = Couvert) = 0$$

D'où

$$\text{Gain(Ciel)} = 0.92 - \left(0.96 * \frac{5}{14} + 0.96 * \frac{5}{14} + 0 * \frac{5}{14}\right)$$
$$= 0.24$$

B. **Calcul de gain de la Température:**

1. Température= Chaude:

$$\begin{cases} C_{Jouer= Oui} = \frac{2}{4} \\ C_{Jouer= Non} = \frac{2}{4} \end{cases} \qquad \frac{|\text{Température= Chaude}|}{|\ C_{Jouer= Oui}, C_{Jouer= Non}\ |} = \frac{4}{14}$$

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

2. <u>Température = Tiède:</u>

$$\begin{cases} C_{Jouer= Oui} = \frac{4}{6} \\ C_{Jouer= Non} = \frac{2}{6} \end{cases} \qquad \frac{\left|\text{ Température= Tiède }\right|}{\left|\ C_{Jouer= Oui}, C_{Jouer= Non}\ \right|} = \frac{6}{14}$$

$$E(Température = Tiède) = -\left(\frac{4}{6}\log_2\left(\frac{4}{6}\right) + \frac{2}{6}\log_2\left(\frac{2}{6}\right)\right) = 0.92$$

3. <u>Température = Fraîche:</u>

$$\begin{cases} C_{Jouer= Oui} = \frac{3}{4} \\ C_{Jouer= Non} = \frac{1}{4} \end{cases} \qquad \frac{\left|\text{ Température= Fraîche }\right|}{\left|\ C_{Jouer= Oui}, C_{Jouer= Non}\ \right|} = \frac{4}{14}$$

$$E(Température = \text{Fraîche}) = -\left(\frac{3}{4}\log_2\left(\frac{3}{4}\right) + \frac{1}{4}\log_2\left(\frac{1}{4}\right)\right) = 0.8$$

D'où

$$\text{Gain(Température)} = 0.92 - (1 * \frac{4}{14} + 0.92 * \frac{6}{14} + 0.8 * \frac{4}{14}) = 0.03$$

C. **Calcule de gain d'Humidité:**

1. <u>Humidité =Elevée:</u>

$$\begin{cases} C_{Jouer= Oui} = \frac{3}{7} \\ C_{Jouer= Non} = \frac{4}{7} \end{cases} \qquad \frac{\left|\text{ Humidité = Elevée }\right|}{\left|\ C_{Jouer= Oui}, C_{Jouer= Non}\ \right|} = \frac{7}{14}$$

$$E(Humidité = Elevée) = -\left(\frac{3}{7}\log_2\left(\frac{3}{7}\right) + \frac{4}{7}\log_2\left(\frac{4}{7}\right)\right) = 0.98$$

2. <u>Humidité =Normale:</u>

$$\begin{cases} C_{Jouer= Oui} = \frac{6}{7} \\ C_{Jouer= Non} = \frac{1}{7} \end{cases} \qquad \frac{\left|\text{ Humidité = Normale }\right|}{\left|\ C_{Jouer= Oui}, C_{Jouer= Non}\ \right|} = \frac{7}{14}$$

$$E(Humidité = \text{Normle}) = -\left(\frac{6}{7}\log_2\left(\frac{6}{7}\right) + \frac{1}{7}\log_2\left(\frac{1}{7}\right)\right) = 0.58$$

D'où

$$\text{Gain(Humidité)} = 0.92 - (0.98 * \frac{7}{14} + 0.58 * \frac{7}{14})$$
$$= 0.14$$

D. **Calcule de gain du Vent:**

1. **<u>Vent =Faible:</u>**

$$\begin{cases} C_{\text{Jouer= Oui}} = \frac{6}{8} \\ C_{\text{Jouer= Non}} = \frac{2}{8} \end{cases} \qquad \frac{\left| \text{ Vent = Faible } \right|}{\left| \ C_{\text{Jouer= Oui}}, C_{\text{Jouer= Non}} \ \right|} = \frac{8}{14}$$

$$E(Vent = faible) = -\left(\frac{6}{8}\log_2\left(\frac{6}{8}\right) + \frac{2}{8}\log_2\left(\frac{2}{8}\right)\right) = 0.8$$

2. **<u>Vent =Fort:</u>**

$$\begin{cases} C_{\text{Jouer= Oui}} = \frac{3}{6} \\ C_{\text{Jouer= Non}} = \frac{3}{6} \end{cases} \qquad \frac{\left| \text{ Vent = Fort } \right|}{\left| \ C_{\text{Jouer= Oui}}, C_{\text{Jouer= Non}} \ \right|} = \frac{8}{14}$$

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

D'où

$$\text{Gain(Vent)} = 0.92 - (0.8 * \frac{8}{14} + 1 * \frac{6}{14})$$
$$= 0.05$$

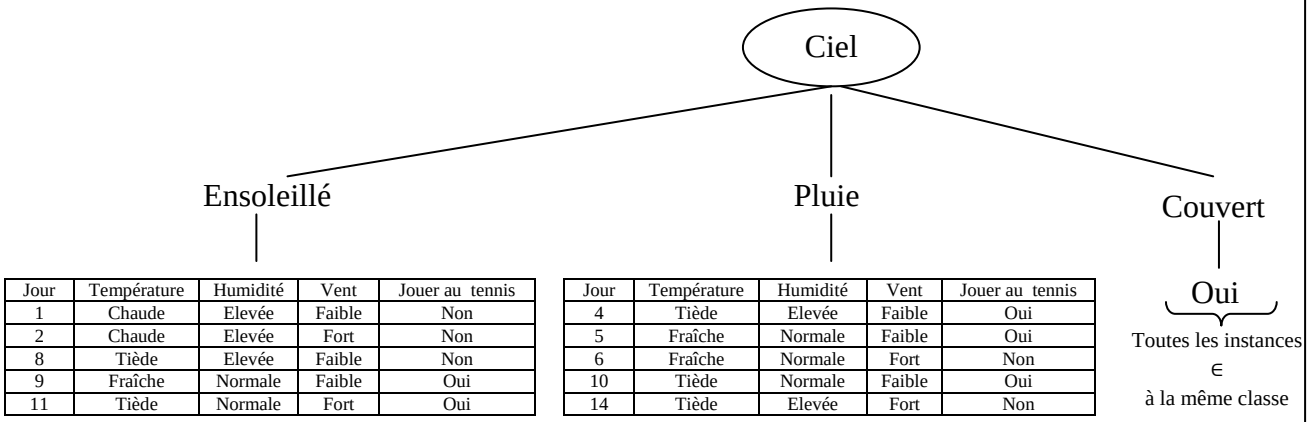
On a :

$$\begin{cases} \text{Gain(Température)} = 0.03 \\ \text{Gain(Ciel)} = 0.24 \\ \text{Gain(Humidité)} = 0.14 \\ \text{Gain(Vent)} = 0.05 \end{cases}$$

Donc :

Gain(Ciel) > Gain(Humidité) >Gain(Vent) >Gain(Température)

Module: Data Mining **Année: 2022-2023**
Nature de document: TD1-Solution
Niveau: L3-STID **N.BERMAD**



Jour	Température	Humidité	Vent	Jouer au tennis
1	Chaude	Elevée	Faible	Non
2	Chaude	Elevée	Fort	Non
8	Tiède	Elevée	Faible	Non
9	Fraîche	Normale	Faible	Oui
11	Tiède	Normale	Fort	Oui

Jour	Température	Humidité	Vent	Jouer au tennis
4	Tiède	Elevée	Faible	Oui
5	Fraîche	Normale	Faible	Oui
6	Fraîche	Normale	Fort	Non
10	Tiède	Normale	Faible	Oui
14	Tiède	Elevée	Fort	Non

- **Raffinement de la branche « Ensoleillé »:**

$E(C_{Jouer= Oui}, C_{Jouer= Non}) = -\left(\frac{2}{5}\log_2\left(\frac{2}{5}\right) + \frac{3}{5}\log_2\left(\frac{3}{5}\right)\right) = 0.96$

Calcule du gain pour chaque attribut:

A. Calcule de gain de la Température:

1. **<u>Température = Chaude:</u>**

$$\begin{cases} C_{Jouer= Oui} = \frac{2}{2} = 0 \\ C_{Jouer= Non} = 1 \end{cases} \qquad \frac{|\ \text{Température= Chaude}\ |}{|\ C_{Jouer= Oui}, C_{Jouer= Non}\ |} = \frac{2}{5}$$

$E(Température = Chaude) = 0$

2. **<u>Température = Tiède:</u>**

$$\begin{cases} C_{Jouer= Oui} = \frac{1}{2} \\ C_{Jouer= Non} = \frac{1}{2} \end{cases} \qquad \frac{|\ \text{Température= Tiède}\ |}{|\ C_{Jouer= Oui}, C_{Jouer= Non}\ |} = \frac{2}{5}$$

$E(Température = Tiède) = 1$

3. **<u>Température = Fraîche:</u>**

$$\begin{cases} C_{Jouer= Oui} = 1 \\ C_{Jouer= Non} = 0 \end{cases} \qquad \frac{|\ \text{Température= Fraîche}\ |}{|\ C_{Jouer= Oui}, C_{Jouer= Non}\ |} = \frac{1}{5}$$

$E(Température = Fraîche) = 0$

D'où

$$\text{Gain(Température)} = 0.96 - (0 * \frac{2}{5} + 1 * \frac{2}{5} + 0 * \frac{1}{5}) = 0.56$$

B. Calcule de gain d'Humidité:

1. Humidité= Elevée:

$$\begin{cases} C_{Jouer=\ Oui} = 0 \\ C_{Jouer=\ Non} = \frac{3}{3} = 1 \end{cases} \qquad \frac{\left|\ \text{Humidité = Elevée}\ \right|}{\left|\ C_{Jouer=\ Oui},\ C_{Jouer=\ Non}\ \right|} = \frac{3}{5}$$

$$E(Humidité = Elevée) = 0$$

2. Humidité= Normale:

$$\begin{cases} C_{Jouer=\ Oui} = \frac{2}{2} = 1 \\ C_{Jouer=\ Non} = 0 \end{cases} \qquad \frac{\left|\ \text{Humidité = Normale}\ \right|}{\left|\ C_{Jouer=\ Oui},\ C_{Jouer=\ Non}\ \right|} = \frac{2}{5}$$

$$E(Humidité = Normale) = 0$$

D'où

$$\text{Gain(Humidité)} = 0.96 - (0 * \frac{3}{5} + 0 * \frac{2}{5})$$
$$= 0.96$$

C. Calcule de gain du Vent:

1. Vent =Faible:

$$\begin{cases} C_{Jouer=\ Oui} = \frac{1}{3} \\ C_{Jouer=\ Non} = \frac{2}{3} \end{cases} \qquad \frac{\left|\ \text{Vent = Faible}\ \right|}{\left|\ C_{Jouer=\ Oui},\ C_{Jouer=\ Non}\ \right|} = \frac{3}{5}$$

$$E(Vent = faible) = -\left(\frac{1}{3}\log_2\left(\frac{1}{3}\right) + \frac{2}{3}\log_2\left(\frac{2}{3}\right)\right) = 0.92$$

2. Vent =Fort:

$$\begin{cases} C_{Jouer=\ Oui} = \frac{1}{2} \\ C_{Jouer=\ Non} = \frac{1}{2} \end{cases} \qquad \frac{\left|\ \text{Vent = Fort}\ \right|}{\left|\ C_{Jouer=\ Oui},\ C_{Jouer=\ Non}\ \right|} = \frac{2}{5}$$

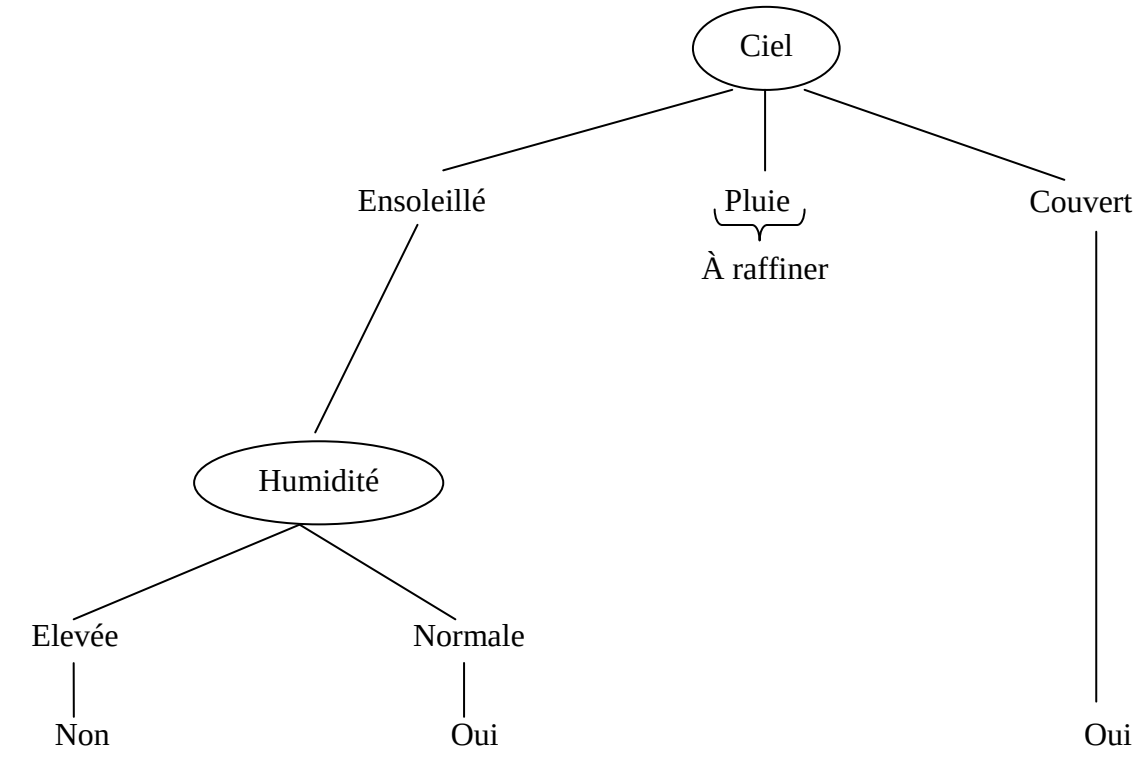
$$E(Vent = Fort) = -\left(\frac{1}{2}\log_2\left(\frac{1}{2}\right) + \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right) = 1$$

D'où

$$\text{Gain(Vent)} = 0.96 - (0.92 * \frac{3}{5} + 1 * \frac{2}{5})$$
$$= 0.008$$

On a :

Gain(Température) = 0.56
Gain(Humidité) = 0.96 ⇨ Gain maximal est de l'humidité
Gain(Vent) = 0.008



- **Raffinement de la branche « Pluie »:**

$E(C_{Jouer= Oui}, C_{Jouer= Non}) = -\left(\frac{3}{5}\log_2\left(\frac{3}{5}\right) + \frac{2}{5}\log_2\left(\frac{2}{5}\right)\right) = 0.96$

Calcul du gain pour chaque attribut:

A. Calcul de gain de la Température :

1. **<u>Température = Tiède:</u>**

$\begin{cases} C_{Jouer= Oui} = \frac{2}{3} \\ C_{Jouer= Non} = \frac{1}{3} \end{cases}$

$$\frac{\left| \text{Température= Tiède} \right|}{\left| C_{Jouer= Oui}, C_{Jouer= Non} \right|} = \frac{3}{5}$$

$$E(Température = Tiède) = 0.92$$

2. **<u>Température = Fraîche:</u>**

$\begin{cases} C_{Jouer= Oui} = \frac{1}{2} \\ C_{Jouer= Non} = \frac{1}{2} \end{cases}$

$$\frac{\left| \text{Température= Fraîche} \right|}{\left| C_{Jouer= Oui}, C_{Jouer= Non} \right|} = \frac{2}{5}$$

$$E(Température = Fraîche) = 1$$

D'où

$$\text{Gain(Température)} = 0.96 - (0.92 * \frac{3}{5} + 1 * \frac{2}{5})$$
$$= 0.01$$

B. Calcule de gain d'Humidité:

1. <u>Humidité= Elevée:</u>

$$\begin{cases} C_{\text{Jouer= Oui}} = \frac{1}{2} \\ C_{\text{Jouer= Non}} = \frac{1}{2} \end{cases} \qquad \frac{\left| \text{Humidité = Elevée} \right|}{\left| C_{\text{Jouer= Oui}}, C_{\text{Jouer= Non}} \right|} = \frac{2}{5}$$

$$E(Humidité = Elevée) = 1$$

2. <u>Humidité= Normale:</u>

$$\begin{cases} C_{\text{Jouer= Oui}} = \frac{2}{3} \\ C_{\text{Jouer= Non}} = \frac{1}{3} \end{cases} \qquad \frac{\left| \text{Humidité = Normale} \right|}{\left| C_{\text{Jouer= Oui}}, C_{\text{Jouer= Non}} \right|} = \frac{3}{5}$$

$$E(Humidité = Normale) = 0.92$$

D'où

$$\text{Gain(Humidité)} = 0.96 - (1 * \frac{2}{5} + 0.92 * \frac{3}{5})$$
$$= 0.008$$

C. Calcule de gain du Vent:

1. <u>Vent =Faible:</u>

$$\begin{cases} C_{\text{Jouer= Oui}} = \frac{3}{3} = 1 \\ C_{\text{Jouer= Non}} = 0 \end{cases} \qquad \frac{\left| \text{Vent = Faible} \right|}{\left| C_{\text{Jouer= Oui}}, C_{\text{Jouer= Non}} \right|} = \frac{3}{5}$$

$$E(Vent = faible) = 0$$

2. <u>Vent =Fort:</u>

$$\begin{cases} C_{\text{Jouer= Oui}} = 0 \\ C_{\text{Jouer= Non}} = \frac{2}{2} = 1 \end{cases} \qquad \frac{\left| \text{Vent = Fort} \right|}{\left| C_{\text{Jouer= Oui}}, C_{\text{Jouer= Non}} \right|} = \frac{2}{5}$$

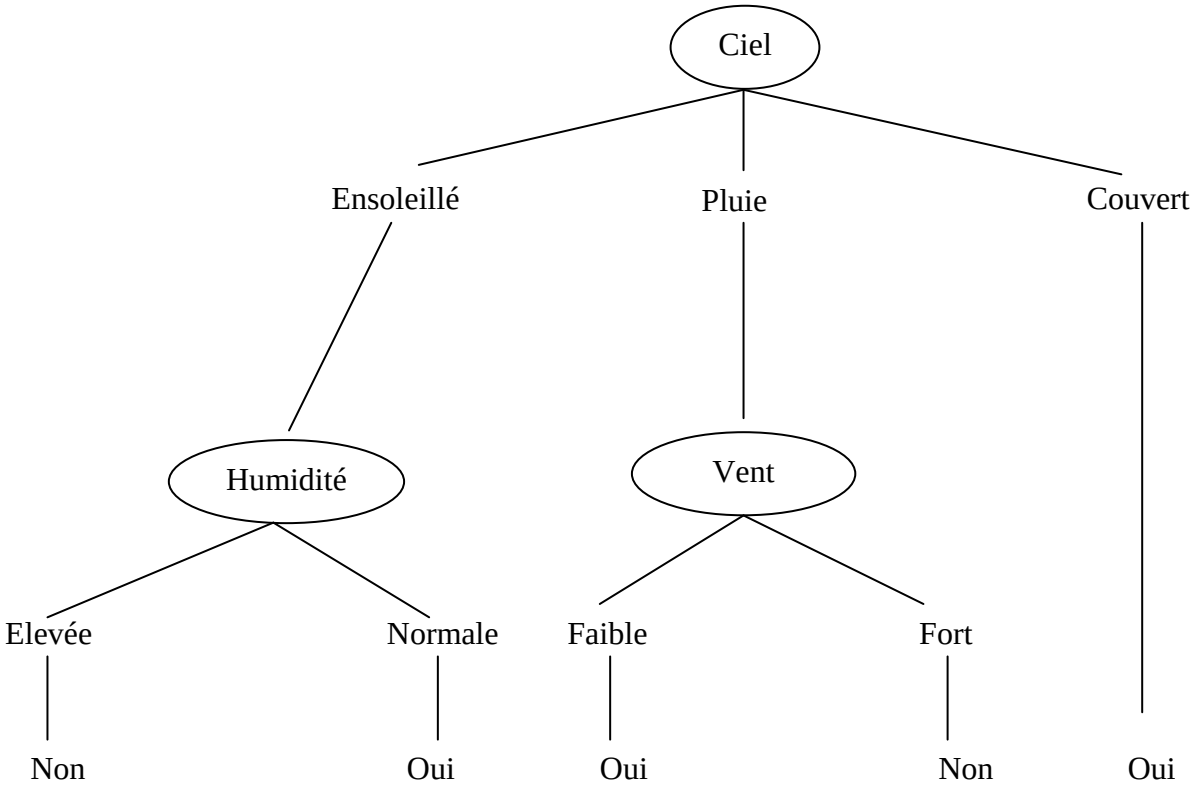
$$E(Vent = Fort) = 0$$

D'où

$$\text{Gain(Vent)} = 0.96 - (0 * \frac{3}{5} + 0 * \frac{2}{5})$$
$$= 0.96$$

On a:

$$\begin{cases} \text{Gain(Température)} = 0.01 \\ \text{Gain(Humidité)} = 0.008 \\ \text{Gain(Vent)} = 0.96 \end{cases} \Rightarrow \text{Vent a un gain maximal}$$



Arbre final

2. Déroulement de l'Algorithme KNN

- Distance euclidienne

$$\begin{cases} d(A,A)=0 \\ d(A,B)=1 \end{cases}$$

- Le paramètre K=4+1=5
- $d(y, x_i)/\ i \in \{1..14\} = \begin{cases} 0 & \text{si } y=x_i \\ 1 & \text{Sinon} \end{cases}$

N°	Ciel	Température	Humidité	Vent	Résultat	Classe
1	0	1	0	0	$\sqrt{\sum_{i=1}^{n}(xi-yi)^2}$ =1	Non
2	0	1	0	1	1.41	Non
3	1	1	0	0	1.41	Oui
4	1	1	0	0	1.41	Oui
5	1	0	1	0	1.41	Oui
6	1	0	1	1	1.73	Non
7	1	0	1	1	1.73	Oui
8	0	1	0	0	1	Non
9	0	0	1	0	1	Oui
10	1	1	1	0	1.73	Oui
11	0	1	1	1	1.73	Oui
12	1	1	0	1	1.73	Oui
13	1	1	1	0	1.73	Oui
14	1	1	0	1	1.73	Non
y	Ensoleillé	Fraîche	Elevée	Faible	Choisir 5 voisins	**Non**

Exercice 4:

1. e est un nouvel individu avec (Ciel=Ensoleillé, Vent = Faible)

   La règle de décision est de classer l'individu e dans la classe jouer au tennis = oui SSI

   P(*Jouer au tennis = Oui|Ciel=Ensoleillé, Vent= Faible***)**
   >
   P(*Jouer au tennis = Non|Ciel=Ensoleillé, Vent = Faible***)**

   En appliquant la règle de Bayes, la règle devient:

   P(*Ciel=Ensoleillé, Vent=Faible | Jouer au tennis = Oui***) P(***Jouer au tennis = Oui***)**
   >
   P(*Ciel=Ensoleillé, Vent=Faible | Jouer au tennis = Non***) P(***Jouer au tennis = Non***)**

   P(*Jouer au tennis = Oui*) $=\frac{9}{14}$

   P(*Jouer au tennis = Non*) $=\frac{5}{14}$

   (*) **P(***Ciel=Ensoleillé, Vent=Faible | Jouer au tennis = Oui***) =**
   P(*Ciel=Ensoleillé | Jouer au tennis = Oui*) P(*Vent=Faible | Jouer au tennis = Oui***)**
   $= \frac{2}{9} * \frac{6}{9} = \frac{12}{81}$
   $= 0.14 * \frac{9}{14} = 0.09$

(**) **P(***Ciel=Ensoleillé, Vent=Faible | Jouer au tennis = Non***) =**
P(*Ciel=Ensoleillé | Jouer au tennis = Non*) P(*Vent=Faible | Jouer au tennis = Non***)**
$= \frac{3}{5} * \frac{2}{5} = \frac{6}{25}$
$= 0.24 + \frac{5}{14} = 0.08$

On a : (*) > (**)
Donc :
e∈ jouer au tennis= oui.

2. e = (Ciel=Ensoleillé, temps= 23°C, Humidité= 70%, Vent=Faible)
(0°C $* \frac{9}{5}$) + 32 = ?? F°
(23°C $* \frac{9}{5}$) + 32 = 73.4 F°

e∈ jouer au tennis= Oui SSI

P(*Ciel=Ensoleillé, Température= 73.4 F°, Humidité= 70%, Vent= Faible |Jouer au tennis = Oui***) P(***Jouer au tennis = Oui***)** (*)
>
P(*Ciel=Ensoleillé, Température= 73.4 F°, Humidité= 70%, Vent= Faible |Jouer au tennis = Non***) P(***Jouer au tennis = Non***)** (**)

(*) ⇔**P(***Ciel= Ensoleillé| Jouer au tennis= Oui***)P(***Température=73.4 F°|Jouer au tennis = Oui***)**
P(*Humidité= 70%|Jouer au tennis= Oui***)P(***Vent=faible |Jouer au tennis = Oui***)**
$= \frac{2}{9} * 0 * \frac{2}{9} * \frac{5}{9} = 0 * \frac{9}{14} = 0$

(**) = 0 ➔ e est un outlier (bruit)

3. **P(***Température =23°C | Jouer au tennis= Oui***)** = 0
4. **P(***60%<Humidité<75%|Jouer au tennis= Oui*)= $\frac{3}{9} =$ 0.33

<u>Exercice 5:</u>

- **Déroulement de l'algorithme C4.5**
- **L'entropie de la population:**

$$\mathbf{E(}C_{Jouer=Oui}, C_{Jouer=Non}) = -\left(\frac{5}{14}\log_2\left(\frac{5}{14}\right) + \frac{9}{14}\log_2\left(\frac{9}{14}\right)\right)$$

= 0.94

A. Calcul de gain d'ensoleillement

1. Ensoleillement = Soleil

$$\begin{cases} C_{Jouer=Oui} = \frac{2}{5} \\ C_{Jouer=Non} = \frac{3}{5} \end{cases} \qquad \frac{|\ \textbf{Soleil}\ |}{|C_{Jouer=Oui}, C_{Jouer=Non}|} = \frac{5}{14}$$

$$E(\text{Ensoleillé}) = -\left(\frac{2}{5}\log_2\left(\frac{2}{5}\right) + \frac{3}{5}\log_2\left(\frac{3}{5}\right)\right) = 0.97$$

2. Ensoleillement = Couvert

$$\begin{cases} C_{Jouer=Oui} = 1 \\ C_{Jouer=Non} = 0 \end{cases} \qquad \frac{|\ \textbf{Couvert}\ |}{|C_{Jouer=Oui}, C_{Jouer=Non}|} = \frac{4}{14}$$

$$E(\text{Couvert}) = 0$$

3. Ensoleillement = Pluie

$$\begin{cases} C_{Jouer=Oui} = \frac{3}{5} \\ C_{Jouer=Non} = \frac{2}{5} \end{cases} \qquad \frac{|\ \textbf{Pluie}\ |}{|C_{Jouer=Oui}, C_{Jouer=Non}|} = \frac{5}{14}$$

$$E(\text{Pluie}) = -\left(\frac{3}{5}\log_2\left(\frac{3}{5}\right) + \frac{2}{5}\log_2\left(\frac{2}{5}\right)\right) = 0.97$$

D'où

$$\text{Gain(Ensoleillement)} = 0.94 - (0.97 * \frac{5}{14} + 0 * \frac{4}{14} + 0.97 * \frac{5}{14})$$
$$= 0.24$$

B. Calcul de gain d'Humidité

- Trier les exemples dans l'ordre croissant (ou décroissant)

Humidité	65	70	70	70	75	78	80	80	80	85	90	90	95	96
N°	8	1	5	11	9	7	10	12	13	3	2	6	4	14
Jouer	Oui	Oui	Oui	Non	Oui	Oui	Non	Oui	Oui	Non	Non	Oui	Non	Oui

Nous coupons entre les exemples:

$$\begin{cases} x_5 \text{ et } x_{11} \rightarrow S_1 = 70 \\ x_{11} \text{ et } x_9 \rightarrow S_2 = 75 \\ x_7 \text{ et } x_{10} \rightarrow S_3 = 80 \\ x_{10} \text{ et } x_{12} \rightarrow S_4 = 80 \\ x_{13} \text{ et } x_3 \rightarrow S_5 = 85 \\ x_2 \text{ et } x_6 \rightarrow S_6 = 90 \\ x_6 \text{ et } x_4 \rightarrow S_7 = 95 \\ x_4 \text{ et } x_{14} \rightarrow S_8 = 96 \end{cases}$$

- E (Humidité<=70)= $-\left(\frac{3}{4}\log_2\left(\frac{3}{4}\right)+\frac{1}{4}\log_2\left(\frac{1}{4}\right)\right)$ $\frac{|\textbf{Humidité <=70}|}{|C_{Jouer=Oui}, C_{Jouer=Non}|}=\frac{4}{14}$

  =0.81

- E (Humidité > 70) $=-\left(\frac{6}{10}\log_2\left(\frac{6}{10}\right)+\frac{4}{10}\log_2\left(\frac{4}{10}\right)\right)$ $\frac{|\textbf{Humidité >70}|}{|C_{Jouer=Oui}, C_{Jouer=Non}|}=\frac{10}{14}$

  =0.97

D'où

Gain(Humidité, S_1= 70) $= 0.94-(0.81*\frac{4}{14}+0.97*\frac{10}{14})$
$= 0.94- 0.92$
$=0.02$

- E (Humidité <= 75) $=-\left(\frac{4}{5}\log_2\left(\frac{4}{5}\right)+\frac{1}{5}\log_2\left(\frac{1}{5}\right)\right)$ $\frac{|\textbf{Humidité <=75}|}{|C_{Jouer=Oui}, C_{Jouer=Non}|}=\frac{5}{14}$

  =0.72

- E (Humidité > 75) $=-\left(\frac{5}{9}\log_2\left(\frac{5}{9}\right)+\frac{4}{9}\log_2\left(\frac{4}{9}\right)\right)$ $\frac{|\textbf{Humidité >75}|}{|C_{Jouer=Oui}, C_{Jouer=Non}|}=\frac{9}{14}$

  =0.99

D'où

$$\text{Gain(Humidité, } S_2 = 75) = 0.94 - (0.72 * \frac{5}{14} + 0.99 * \frac{9}{14})$$
$$= 0.94 - 0.89$$
$$= 0.05$$

- E (Humidité <= 80) $= -\left(\frac{7}{9}\log_2\left(\frac{7}{9}\right) + \frac{2}{9}\log_2\left(\frac{2}{9}\right)\right)$ $\frac{|\textbf{Humidité <= 80}|}{|C_{\text{Jouer=Oui}}, C_{\text{Jouer=Non}}|} = \frac{9}{14}$

$$= 0.76$$

- E (Humidité >80) $= -\left(\frac{2}{5}\log_2\left(\frac{2}{5}\right) + \frac{3}{5}\log_2\left(\frac{3}{5}\right)\right)$ $\frac{|\textbf{Humidité > 80}|}{|C_{\text{Jouer=Oui}}, C_{\text{Jouer=Non}}|} = \frac{5}{14}$

$$= 0.97$$

D'où

$$\text{Gain(Humidité, } S_3 = 80) = 0.94 - (0.76 * \frac{9}{14} + 0.97 * \frac{5}{14})$$
$$= 0.94 - 0.83$$
$$= 0.11$$

- E (Humidité <= 85) $= -\left(\frac{7}{10}\log_2\left(\frac{7}{10}\right) + \frac{3}{10}\log_2\left(\frac{3}{10}\right)\right)$ $\frac{|\textbf{Humidité <= 85}|}{|C_{\text{Jouer=Oui}}, C_{\text{Jouer=Non}}|} = \frac{10}{14}$

$$= 0.88$$

- E (Humidité >85) $= -\left(\frac{2}{4}\log_2\left(\frac{2}{4}\right) + \frac{2}{4}\log_2\left(\frac{2}{4}\right)\right)$ $\frac{|\textbf{Humidité > 85}|}{|C_{\text{Jouer=Oui}}, C_{\text{Jouer=Non}}|} = \frac{4}{14}$

$$= 1$$

D'où

$$\text{Gain(Humidité, } S_5 = 85) = 0.94 - (0.88 * \frac{10}{14} + 1 * \frac{4}{14})$$
$$= 0.94 - 0.91$$
$$= 0.03$$

- E (Humidité <= 90) $= -\left(\frac{8}{12}\log_2\left(\frac{8}{12}\right) + \frac{4}{12}\log_2\left(\frac{4}{12}\right)\right)$ $\frac{|\textbf{Humidité <= 90}|}{|C_{\text{Jouer=Oui}}, C_{\text{Jouer=Non}}|} = \frac{12}{14}$

$$= 0.92$$

- E (Humidité >90) $= -\left(\frac{1}{2}\log_2\left(\frac{1}{2}\right) + \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right)$

$$= 1$$

$$\frac{\left|\textbf{Humidité} > \textbf{90}\right|}{\left|C_{\text{Jouer=Oui}}, C_{\text{Jouer=Non}}\right|} = \frac{2}{14}$$

D'où

- Gain(Humidité, S_6= 90) $= 0.94 - (0.92 * \frac{12}{14} + 1 * \frac{2}{14})$
  $= 0.94 - 0.92$
  $=0.02$

- Gain(Humidité, S_7 = 95) $= 0.94 - (0.96 * \frac{13}{14} + 0 * \frac{1}{14})$
  $= 0.94 - 0.89$
  $=0.05$

- Gain(Humidité, S_8= 96) $= 0.94 - (0.94 * 1 + 0 * 0)$
  $= 0.94 - 0.94$
  $=0$

Le gain maximal d'humidité et avec le seuil S3=80

C. Calcul de gain de la Température

Trier les exemples dans l'ordre croissant

Température	64	65	68	69	70	71	72	72	75	75	80	81	83	85
N°	8	11	13	5	14	10	4	6	1	12	2	9	7	3
Jouer	Oui	Non	Oui	Oui	Oui	Non	Non	Oui	Oui	Oui	Non	Oui	Oui	Non

Nous coupons entre les exemples :

x_8 et x_{11} $\xrightarrow{S_1}$ $S_1 = 65$
x_{11} et x_{13} $\xrightarrow{S_2}$ $S_2 = 68$
x_{14} et x_{10} $\xrightarrow{S_3}$ $S_3 = 71$
x_4 et x_6 $\xrightarrow{S_4}$ $S_4 = 72$
x_{12} et x_2 $\xrightarrow{S_5}$ $S_5 = 80$
x_2 et x_9 $\xrightarrow{S_6}$ $S_6 = 81$
x_7 et x_3 $\xrightarrow{S_7}$ $S_7 = 85$

- E (Température<= 65) $= -\left(\frac{1}{2}\log_2\left(\frac{1}{2}\right) + \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right)$ $\frac{\left|\textbf{Température <=65}\right|}{\left|C_{Jouer=Oui}, C_{Jouer=Non}\right|} = \frac{2}{14}$

  $= 1$

- E (Température>65) $= -\left(\frac{8}{12}\log_2\left(\frac{8}{12}\right) + \frac{4}{12}\log_2\left(\frac{4}{12}\right)\right)$ $\frac{\left|\textbf{Température >65}\right|}{\left|C_{Jouer=Oui}, C_{Jouer=Non}\right|} = \frac{12}{14}$

  $=0.92$

D'où

$$\text{Gain(Température, } S_1 = 65) = 0.94 - \left(1 * \frac{2}{14} + 0.92 * \frac{12}{14}\right)$$
$$= 0$$

- E (Température<= 68) $= -\left(\frac{2}{3}\log_2\left(\frac{2}{3}\right) + \frac{1}{3}\log_2\left(\frac{1}{3}\right)\right)$ $\frac{\left|\textbf{Température <=68}\right|}{\left|C_{Jouer=Oui}, C_{Jouer=Non}\right|} = \frac{3}{14}$

  $= 0.92$

- E (Température>68) $= -\left(\frac{7}{11}\log_2\left(\frac{7}{11}\right) + \frac{4}{11}\log_2\left(\frac{4}{11}\right)\right)$ $\frac{\left|\textbf{Température >68}\right|}{\left|C_{Jouer=Oui}, C_{Jouer=Non}\right|} = \frac{11}{14}$

  $=0.93$

D'où

$$\text{Gain(Température, } S_2 = 68) = 0.94 - \left(0.92 * \frac{3}{14} + 0.93 * \frac{11}{14}\right)$$
$$= 0.03$$

- E (Température<= 71) $= -\left(\frac{4}{6}\log_2\left(\frac{4}{6}\right) + \frac{2}{6}\log_2\left(\frac{2}{6}\right)\right)$ $\frac{\left|\textbf{Température <=71}\right|}{\left|C_{Jouer=Oui}, C_{Jouer=Non}\right|} = \frac{6}{14}$

  $= 0.92$

- E (Température>71) $= -\left(\frac{5}{8}\log_2\left(\frac{5}{8}\right) + \frac{3}{8}\log_2\left(\frac{3}{8}\right)\right)$ $\frac{\left|\textbf{Température >71}\right|}{\left|C_{Jouer=Oui}, C_{Jouer=Non}\right|} = \frac{8}{14}$

  $=0.95$

D'où

$$\text{Gain(Température, } S_3 = 71) = 0.94 - (0.92 * \frac{6}{14} + 0.95 * \frac{8}{14})$$
$$= 0.01$$

- E (Température<= 72) $= -\left(\frac{5}{8}\log_2\left(\frac{5}{8}\right) + \frac{3}{8}\log_2\left(\frac{3}{8}\right)\right)$ $\quad \frac{|\textbf{Température <=72}|}{|C_{Jouer=Oui}, C_{Jouer=Non}|} = \frac{8}{14}$

$$= 0.95$$

- E (Température>72) $= -\left(\frac{4}{6}\log_2\left(\frac{4}{6}\right) + \frac{2}{6}\log_2\left(\frac{2}{6}\right)\right)$ $\quad \frac{|\textbf{Température >72}|}{|C_{Jouer=Oui}, C_{Jouer=Non}|} = \frac{6}{14}$

$$=0.92$$

D'où

$$\text{Gain(Température, } S_4 = 72) = 0.94 - (0.95 * \frac{8}{14} + 0.92 * \frac{6}{14})$$
$$= 0.01$$

- E (Température<= 80) $= -\left(\frac{7}{11}\log_2\left(\frac{7}{11}\right) + \frac{4}{11}\log_2\left(\frac{4}{11}\right)\right)$ $\quad \frac{|\textbf{Température <=80}|}{|C_{Jouer=Oui}, C_{Jouer=Non}|} = \frac{11}{14}$

$$= 0.93$$

- E (Température>80) $= -\left(\frac{2}{3}\log_2\left(\frac{2}{3}\right) + \frac{1}{3}\log_2\left(\frac{1}{3}\right)\right)$ $\quad \frac{|\textbf{Température >80}|}{|C_{Jouer=Oui}, C_{Jouer=Non}|} = \frac{3}{14}$

$$=0.92$$

D'où

$$\text{Gain(Température, } S_5 = 80) = 0.94 - (0.93 * \frac{11}{14} + 0.92 * \frac{3}{14})$$
$$= 0.03$$

- E (Température<= 81) $= -\left(\frac{8}{12}\log_2\left(\frac{8}{12}\right) + \frac{4}{12}\log_2\left(\frac{4}{12}\right)\right)$ $\quad \frac{|\textbf{Température <=81}|}{|C_{Jouer=Oui}, C_{Jouer=Non}|} = \frac{12}{14}$

$$= 0.92$$

- E (Température>81) $= -\left(\frac{1}{2}\log_2\left(\frac{1}{2}\right) + \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right)$

$$\frac{\left|\textbf{Température >81}\right|}{\left|C_{\text{Jouer=Oui}}, C_{\text{Jouer=Non}}\right|} = \frac{2}{14}$$

$=1$

D'où

$$\text{Gain(Température, } S_6 = 81) = 0.94 - \left(1 * \frac{2}{14} + 0.92 * \frac{12}{14}\right)$$
$$= 0.01$$

- E (Température<= 85) $= -\left(\frac{9}{14}\log_2\left(\frac{9}{14}\right) + \frac{5}{14}\log_2\left(\frac{5}{14}\right)\right)$

$$\frac{\left|\textbf{Température <=85}\right|}{\left|C_{\text{Jouer=Oui}}, C_{\text{Jouer=Non}}\right|} = 1$$

$= 0.94$

- E (Température>85) $= 0$

$$\frac{\left|\textbf{Température >85}\right|}{\left|C_{\text{Jouer=Oui}}, C_{\text{Jouer=Non}}\right|} = 0$$

D'où

$$\text{Gain(Température, } S_7 = 80) = 0.94 - \left(0 * 0 + 0.94 * \frac{14}{14}\right)$$
$$=00$$

D. Calcul de gain du Vent

1. Vent = Oui

$$\begin{cases} C_{\text{Jouer=Oui}} = \frac{3}{6} \\ C_{\text{Jouer=Non}} = \frac{3}{6} \end{cases}$$

$$\frac{\left|\textbf{Vent = Oui}\right|}{\left|C_{\text{Jouer=Oui}}, C_{\text{Jouer=Non}}\right|} = \frac{6}{14}$$

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

2. Vent = Non

$$\begin{cases} C_{\text{Jouer=Oui}} = \frac{6}{8} \\ C_{\text{Jouer=Non}} = \frac{2}{8} \end{cases}$$

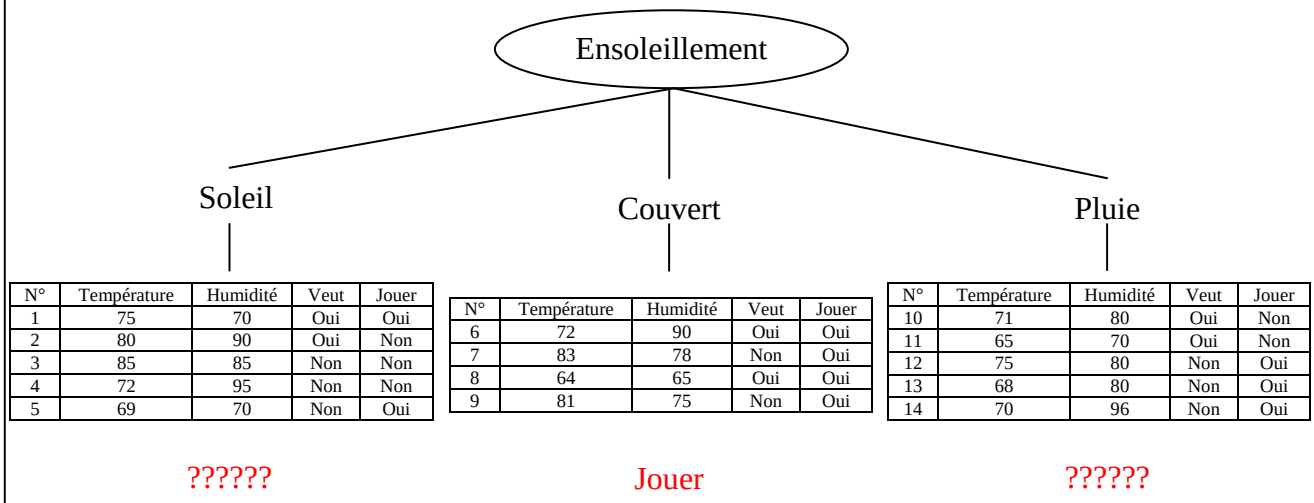
$$\frac{\left|\textbf{Vent = Non}\right|}{\left|C_{\text{Jouer=Oui}}, C_{\text{Jouer=Non}}\right|} = \frac{8}{14}$$

$$E(\text{Vent} = \text{Non}) = -\left(\frac{6}{8}\log_2\left(\frac{6}{8}\right) + \frac{2}{8}\log_2\left(\frac{2}{8}\right)\right) = 0.81$$

D'où

$$\text{Gain(Vent)} = 0.94 - (1 * \frac{6}{14} + 0.81 * \frac{8}{14})$$
$$= 0.05$$

L'attribut ensoleillement a la plus grande valeur du gain en information. Donc, c'est le nœud racine de l'arbre:



N°	Température	Humidité	Veut	Jouer
1	75	70	Oui	Oui
2	80	90	Oui	Non
3	85	85	Non	Non
4	72	95	Non	Non
5	69	70	Non	Oui

N°	Température	Humidité	Veut	Jouer
6	72	90	Oui	Oui
7	83	78	Non	Oui
8	64	65	Oui	Oui
9	81	75	Non	Oui

N°	Température	Humidité	Veut	Jouer
10	71	80	Oui	Non
11	65	70	Oui	Non
12	75	80	Non	Oui
13	68	80	Non	Oui
14	70	96	Non	Oui

?????? Jouer ??????

- **Raffinement de la branche « Soleil »:**

L'entropie de la population:

$$E(C_{Jouer=Oui}, C_{Jouer=Non}) = -\left(\frac{2}{5}\log_2\left(\frac{2}{5}\right) + \frac{3}{5}\log_2\left(\frac{3}{5}\right)\right)$$
$$= \mathbf{0.97}$$

A. Calcul de gain de la Température :

Température	69	72	75	80	85
N°	5	4	1	2	3
Jouer	Oui	Non	Oui	Oui	Non

Nous coupons entre les exemples :

x_5 et x_4 ⟶ $S_1 = 72$
x_4 et x_1 ⟶ $S_2 = 75$
x_2 et x_3 ⟶ $S_3 = 85$

- E (Température<= 72) $= -\left(\frac{1}{2}\log_2\left(\frac{1}{2}\right) + \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right)$

  $= 1$

  $\frac{|\textbf{Température <=72}|}{|C_{Jouer=Oui}, C_{Jouer=Non}|} = \frac{2}{5}$

- E (Température>72) $= -\left(\frac{2}{3}\log_2\left(\frac{2}{3}\right) + \frac{1}{3}\log_2\left(\frac{1}{3}\right)\right)$

  $=0.92$

  $\frac{|\textbf{Température >72}|}{|C_{Jouer=Oui}, C_{Jouer=Non}|} = \frac{3}{5}$

D'où

$$\text{Gain(Température, } S_1 = 72) = 0.97 - (1 * \frac{2}{5} + 0.92 * \frac{3}{5})$$
$$= 0.018$$

- E (Température<= 75) $= -\left(\frac{2}{3}\log_2\left(\frac{2}{3}\right) + \frac{1}{3}\log_2\left(\frac{1}{3}\right)\right)$

  $= 0.92$

  $\frac{|\textbf{Température <=75}|}{|C_{Jouer=Oui}, C_{Jouer=Non}|} = \frac{3}{5}$

- E (Température>75) $= -\left(\frac{1}{2}\log_2\left(\frac{1}{2}\right) + \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right)$

  $= 1$

  $\frac{|\textbf{Température >75}|}{|C_{Jouer=Oui}, C_{Jouer=Non}|} = \frac{2}{5}$

D'où

$$\text{Gain(Température, } S_2 = 75) = 0.97 - (0.92 * \frac{3}{5} + 1 * \frac{2}{5})$$
$$= 0.018$$

- E (Température<= 85) $= -\left(\frac{3}{5}\log_2\left(\frac{3}{5}\right) + \frac{2}{5}\log_2\left(\frac{2}{5}\right)\right)$

  $= 0.97$

  $\frac{|\textbf{Température <=85}|}{|C_{Jouer=Oui}, C_{Jouer=Non}|} = 1$

- E (Température>85) $= 0$

  $\frac{|\textbf{Température >85}|}{|C_{Jouer=Oui}, C_{Jouer=Non}|} = 0$

D'où

$$\text{Gain(Température, } S_3 = 85) = 0.97 - (0.97 * 1 + 0 * 0)$$
$$=0$$

B. Calcul de gain d'Humidité:

Humidité	70	70	85	90	95
N°	1	5	3	2	4
Jouer	Oui	Oui	Non	Non	Non

Nous coupons entre les exemples: x_5 et x_3 ➔S= $\frac{85+70}{2} = 77.5$ (Car l'écart est grand)

- E (Humidité <= 77.5) $= -\left(\frac{2}{2}\log_2\left(\frac{2}{2}\right) + 0\log_2(0)\right)$

  $= 0$

  $\frac{\left|\textbf{Humidité <=77.5}\right|}{\left|C_{\text{Jouer=Oui}}, C_{\text{Jouer=Non}}\right|} = \frac{2}{5}$

- E (Humidité >77.5) $= -(0\ \log_2(0) + 1\log_2(1))$

  $=0$

  $\frac{\left|\textbf{Humidité>77.5}\right|}{\left|C_{\text{Jouer=Oui}}, C_{\text{Jouer=Non}}\right|} = \frac{3}{5}$

D'où

Gain(Humidité, S= 77.5) $= 0.97 - (0 * \frac{2}{5} + 0 * \frac{3}{5})$

$= 0.97$

C. Calcul de gain du Vent

1. Vent = Oui

$$\begin{cases} C_{\text{Jouer=Oui}} = \frac{1}{2} \\ C_{\text{Jouer=Non}} = \frac{1}{2} \end{cases} \qquad \frac{\left|\textbf{Vent = Oui}\right|}{\left|C_{\text{Jouer=Oui}}, C_{\text{Jouer=Non}}\right|} = \frac{2}{5}$$

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

2. Vent = Non

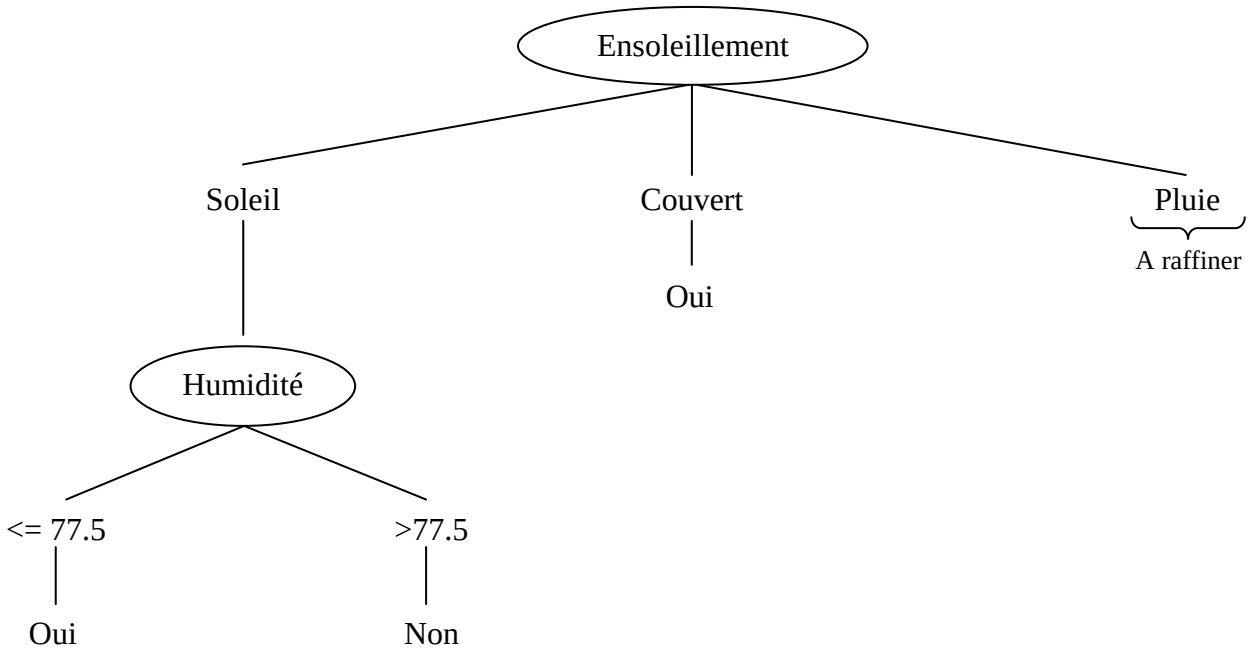
$$\begin{cases} C_{\text{Jouer=Oui}} = \frac{1}{3} \\ C_{\text{Jouer=Non}} = \frac{2}{3} \end{cases} \qquad \frac{\left|\textbf{Veut = Non}\right|}{\left|C_{\text{Jouer=Oui}}, C_{\text{Jouer=Non}}\right|} = \frac{3}{5}$$

$$E(\text{Vent} = \text{Non}) = -\left(\frac{1}{3}\log_2\left(\frac{1}{3}\right) + \frac{2}{3}\log_2\left(\frac{2}{3}\right)\right) = 0.92$$

D'où

$$\text{Gain(Vent)} = 0.97 - (1 * \frac{2}{5} + 0.92 * \frac{3}{5})$$
$$= 0.018$$

Nous choisissons l'attribut humidité ayant un gain maximal. Donc, l'arbre de décision est:



- **Raffinement de la branche « Pluie »:**

L'entropie de la population:

$$E(C_{Jouer=Oui}, C_{Jouer=Non}) = -\left(\frac{3}{5}\log_2\left(\frac{3}{5}\right) + \frac{2}{5}\log_2\left(\frac{2}{5}\right)\right)$$
$$= \mathbf{0.97}$$

A. **Calcul de gain de la Température :**

Température	65	68	70	71	75
N°	11	13	14	10	12
Jouer	Non	Oui	Oui	Non	Oui

Nous coupons entre:

x_{11} et x_{13} ⟶ $S_1 = 68$
x_{14} et x_{10} ⟶ $S_2 = 71$
x_{10} et x_{12} ⟶ $S_3 = 75$

- E (Température<= 68) $= -\left(\frac{1}{2}\log_2\left(\frac{1}{2}\right) + \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right)$

  $= 1$

  $\frac{|\textbf{Température <=68}|}{|C_{Jouer=Oui}, C_{Jouer=Non}|} = \frac{2}{5}$

- E (Température>68) $= -\left(\frac{2}{3}\log_2\left(\frac{2}{3}\right) + \frac{1}{3}\log_2\left(\frac{1}{3}\right)\right)$

  $=0.92$

  $\frac{|\textbf{Température >68}|}{|C_{Jouer=Oui}, C_{Jouer=Non}|} = \frac{3}{5}$

D'où

$\text{Gain(Température, } S_1 = 68) = 0.97 - (1 * \frac{2}{5} + 0.92 * \frac{3}{5})$

$= 0.018$

- E (Température<= 71) $= -\left(\frac{2}{4}\log_2\left(\frac{2}{4}\right) + \frac{2}{4}\log_2\left(\frac{2}{4}\right)\right)$

  $= 1$

  $\frac{|\textbf{Température <=71}|}{|C_{Jouer=Oui}, C_{Jouer=Non}|} = \frac{4}{5}$

- E (Température>71) $= -(1\ \log_2(1) + 0\log_2(0))$

  $= 0$

  $\frac{|\textbf{Température >71}|}{|C_{Jouer=Oui}, C_{Jouer=Non}|} = \frac{1}{5}$

D'où

$\text{Gain(Température, } S_2 = 71) = 0.97 - (1 * \frac{4}{5} + 0 * \frac{1}{5})$

$= 0.17$

- E (Température<= 75) $= -\left(\frac{3}{5}\log_2\left(\frac{3}{5}\right) + \frac{2}{5}\log_2\left(\frac{2}{5}\right)\right)$

  $= 0.97$

  $\frac{|\textbf{Température <=75}|}{|C_{Jouer=Oui}, C_{Jouer=Non}|} = \frac{5}{5} = 1$

- E (Température>75) $= 0$

  $\frac{|\textbf{Température >75}|}{|C_{Jouer=Oui}, C_{Jouer=Non}|} = 0$

D'où

$\text{Gain(Température, } S_3 = 75) = 0.97 - (0.97 * 1 + 0 * 0)$

$= 0$

B. **Calcul de gain d'humidité:**

Humidité	70	80	80	80	96
N°	11	10	12	13	14
Jouer	Non	Non	Oui	Oui	Oui

Nous coupons entre x_{10}et x_{12} ➔ S= 80

- E (Humidité<= 80) $= -\left(\frac{2}{4}\log_2\left(\frac{2}{4}\right) + \frac{2}{4}\log_2\left(\frac{2}{4}\right)\right)$ $\quad \frac{\left|\textbf{ Humidité <= 80 }\right|}{\left|C_{\text{Jouer=Oui}}, C_{\text{Jouer=Non}}\right|} = \frac{4}{5}$

  =1

- E (Humidité>80) $= -\left(\frac{1}{1}\log_2(1) + 0\log_2(0)\right)$ $\quad \frac{\left|\textbf{ Humidité> 80 }\right|}{\left|C_{\text{Jouer=Oui}}, C_{\text{Jouer=Non}}\right|} = \frac{1}{5}$

  =0

D'où

Gain(Humidité, S= 80) = $0.97 - (1 * \frac{4}{5} + 0 * \frac{1}{5})$

= 0.17

C. **Calcul de gain du Vent**

1. **Vent = Oui**

$$\begin{cases} C_{\text{Jouer=Oui}} = \frac{0}{2} = 0 \\ C_{\text{Jouer=Non}} = \frac{2}{2} = 1 \end{cases} \qquad \frac{\left|\textbf{Vent = Oui }\right|}{\left|C_{\text{Jouer=Oui}}, C_{\text{Jouer=Non}}\right|} = \frac{2}{5}$$

$$E(\text{Vent} = \text{Oui}) = -(0\ \log_2(0) + 1\log_2(1)) = 0$$

2. **Vent = Non**

$$\begin{cases} C_{\text{Jouer=Oui}} = \frac{3}{3} = 1 \\ C_{\text{Jouer=Non}} = 0 \end{cases} \qquad \frac{\left|\textbf{Vent = Non }\right|}{\left|C_{\text{Jouer=Oui}}, C_{\text{Jouer=Non}}\right|} = \frac{3}{5}$$

$$E(\text{Vent} = \text{Non}) = -(1\ \log_2(1) + 0\log_2(0)) = 0$$

D'où

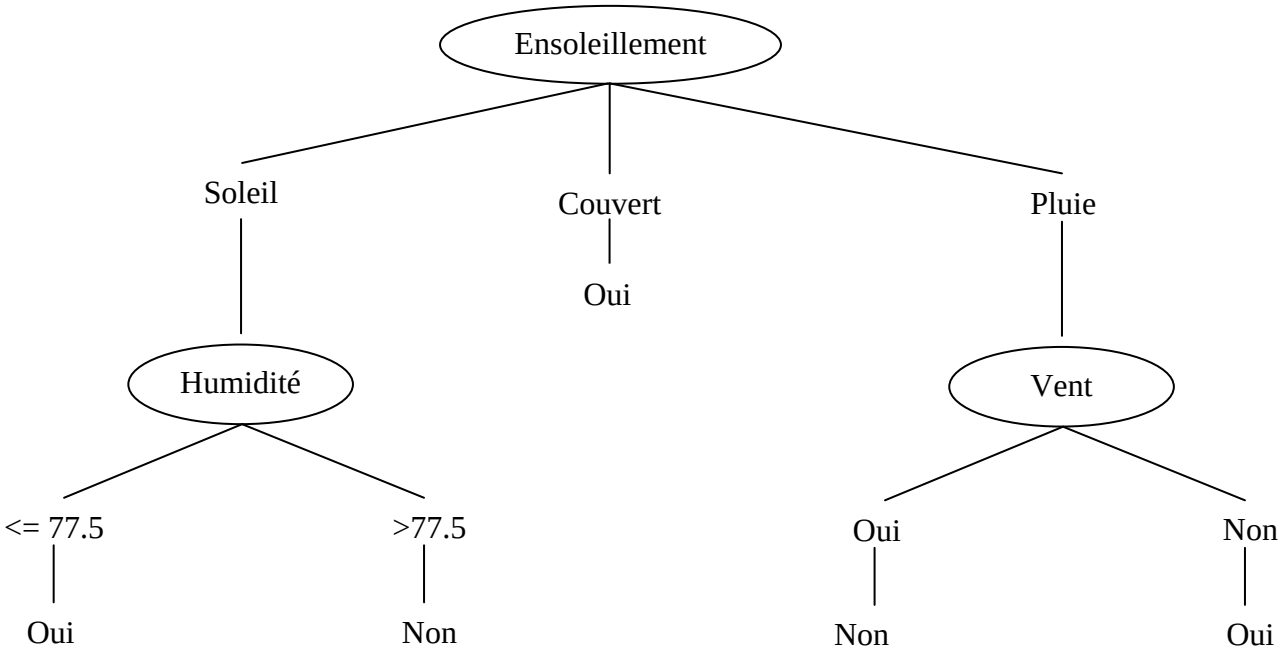
Gain(Vent) = $0.97 - (0 * \frac{3}{5} + 0 * \frac{2}{5})$

= 0.97

Débriefing : Nous avons:

Gain (Vent) = 0.97
Gain (Humidité, S) = 0.17 ➔ nous choisissons vent ayant un gain maximal
Gain (Température, S_2) = 0.17

L'arbre final est:



3. Ici la précision est calculée sur l'ensemble d'apprentissage et non pas de test. Pour cela, nous devons construire la matrice de confusion:

Classe actuelle	Classe prédite	
	Jouer=Oui	Jouer=Non
Jouer=Oui	F11=9	F10=00
Jouer=Non	F01=00	F00 =5

Nb: F_{ij}=c'est le nombre des instances de la classe i prédite dans la classe j.

La prédiction de la classe des données d'apprentissages avec lesquelles on a construit l'arbre de décision.

Donc :

$$\text{Précision} = \frac{\text{Nombre des prédictions correctes}}{\text{Nombre totale des prédiction}}$$

$$= \frac{\text{F11} + \text{F00}}{\text{F11} + \text{F10} + \text{F01} + \text{F00}}$$

$$= \frac{9 + 5}{9 + 0 + 0 + 5}$$

$$= \frac{14}{14}$$

$$= (1 * 100)\%$$

$$= 100\%$$

Une précision de 100% sur l'ensemble d'apprentissage illustre bien le phénomène de sur apprentissage qui ne signifie pas forcément que le modèle obtenu généralise bien sur des nouvelles données.