

### TD5 : Construction d'arbre de Décisions Algorithmes ID3 et C4.5

**Vous êtes encouragés à utiliser l'internet pour les questions de cours, pas les exercices.**

**Questions :**

1. Avantages et limites des arbres de décisions ?
2. Faire un tableau comparatif des algorithmes de classification supervisée vus aux cours.
3. Domaines d'applications ?
4. Quel est le proverbe arabe sur lequel s'appuie la construction des arbres de décisions?

**Exercice 1 :** Pour le tableau Data Weather « symbolique » vu au cours, construire l'arbre de décision, en utilisant le **Rapport de Gain**. Sans consulter les valeurs données au cours.

Exercice 2 : En utilisant ID3, la **mesure du gain**, et la table d'apprentissage suivante (classe maladie), donner l'arbre de décision (détailler les étapes, et les calculs). Classifier l'instance : oui Abdomen non ?

Fièvre	Douleur	Toux	Maladie
oui	Abdomen	non	Appendicite
non	Abdomen	oui	Appendicite
oui	gorge	non	rhume
oui	gorge	oui	rhume
non	gorge	oui	mal de gorge
oui	non	non	aucune
oui	non	oui	rhume
non	non	oui	refroidissement
non	non	non	aucune

**Exercice 2 :** En utilisant ID3, la **mesure du gain**, et la table d'apprentissage suivante (classe species), donner l'arbre de décision (détailler les étapes, et les calculs)

Toothed	Hair	Breathes	Legs	species
Toothed	Hair	Breathes	Legs	Mammal
Toothed	Hair	Breathes	Legs	Mammal
Toothed	Not Hair	Breathes	Not Legs	Reptile
Not Toothed	Hair	Breathes	Legs	Mammal
Toothed	Hair	Breathes	Legs	Mammal
Toothed	Hair	Breathes	Legs	Mammal
Toothed	Not Hair	Not Breathes	Not Legs	Reptile
Toothed	Not Hair	Breathes	Not Legs	Reptile
Toothed	Not Hair	Breathes	Legs	Mammal
Not Toothed	Not Hair	Breathes	Legs	Reptile

**Exercice 3 :** En considérant seulement l'attribut Taille pour la classe Sexe. Trouver le meilleur point de split (S) pour l'algorithme C4.5 Pour le pseudo-attribut (Taille <S)

Taille	161	185	184	164	169	179	176	175	180
Sexe	F	F	M	F	M	F	F	M	M

**EXERCICE N°4 :** Expliquer le principe de **C4.5**.

Qu'a-t-il apporté de plus qu'ID3 ?

En utilisant l'algorithme C4.5 et la mesure **rapport de gain**, sur la table d'apprentissage suivantes,

Donner le nœud racine de l'arbre

**Les calculs doivent être détaillés**

<u>A</u>	<u>B</u>	<u>C</u>	<u>Class</u> <u>e</u>	
<b>S</b>	8 5	F	Y	
<b>S</b>	9 0	T	N	
<b>O</b>	8 6	F	Y	
<b>R</b>	8 0	F	Y	
<b>R</b>	7 0	T	Y	
<b>O</b>	6 5	T	Y	
<b>S</b>	9 5	F	N	
<b>S</b>	7 0	F	N	

**Base d'Apprentissage**

**Exercice 5 :** En utilisant l'algorithme ID3 et la mesure du rapport du gain et la base d'apprentissage suivante, construire l'arbre de décision. **Montrer toutes les étapes du calcul. Dessiner le ou les arbres possibles.**

**Trouver la classe de l'instance :** F M M

<b>E</b>	<b>C</b>	<b>R</b>	Classe
M	E	E	Y
F	F	M	N
E	M	F	N
M	M	M	Y
F	M	E	Y
E	E	F	Y
F	F	F	N
M	M	E	Y

**Exercice 6 :**

En utilisant l'algorithme ID3 et la mesure du gain, construire l'arbre du tableau d'apprentissage, puis classer l'instance :  $\leq 30$  low no excellent

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31...40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31...40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31...40	medium	no	excellent	yes
31...40	high	yes	fair	yes
>40	medium	no	excellent	no