

## Arbres de décision

**Exercice 1** Soit les exemples suivants ayant trois attributs et appartenant à deux classes :

N°	Att1	Att2	Att3	Classe	N°	Att1	Att2	Att3	Classe
1	A	L	T	C1	8	B	L	T	C1
2	A	H	T	C2	9	B	H	F	C1
3	A	H	F	C2	10	C	H	T	C2
4	A	H	F	C2	11	C	L	T	C2
5	A	L	F	C1	12	C	H	F	C1
6	B	H	T	C1	13	C	H	F	C1
7	B	H	F	C1	14	C	H	F	C1

En utilisant l'algorithme ID3 :

1. Calculer l'entropie de l'ensemble d'exemples par rapport à la valeur de la classe.
2. Construire l'arbre de décision appris de cette table.
3. Donner la précision de l'arbre appris sur l'ensemble d'entraînement.
4. Reconstruire l'arbre en utilisant le Gini Index.

**Exercice 2** Une région souhaiterait comprendre les différences d'investissements dans plusieurs villes et villages. Un relevé des niveaux d'investissement a été réalisé en prenant en compte plusieurs types de critères :

- présence d'une gare : oui/non ;
- statut : ville/village ;
- distance par rapport à la capitale : proche/loin ;
- niveau d'investissement (classe) : faible/moyen/élevé.

Etablir, en utilisant l'algorithme ID3 un arbre de décision, expliquant les niveaux d'investissement, à partir des données suivantes :

Gare	Statut	Distance	Investissement
non	village	proche	moyen
oui	ville	proche	élevé
non	village	loin	faible
oui	village	proche	élevé
non	ville	loin	faible
oui	ville	loin	moyen
non	ville	proche	élevé

**Exercice 3** Dans un hôpital, on souhaite construire un arbre de décision pour la prédiction du risque des patients d'avoir une certaine maladie en fonction de leur age et de deux symptômes booléens (vrai ou faux) appelés S1 et S2. Le risque est évalué selon trois valeurs F (faible), M (moyen) et E (élevé), l'age est discrétisé selon trois classes (jeune, adulte et senior).

L'hôpital dispose de la table suivante :

N°	Age	S1	S2	Risque	N°	Age	S1	S2	Risque
1	Jeune	F	V	F	6	Jeune	F	F	F
2	Jeune	V	V	E	7	Adulte	V	F	M
3	Adulte	F	F	F	8	Adulte	V	V	M
4	Senior	V	F	E	9	Senior	F	F	F
5	Senior	F	V	M	10	Senior	V	V	E

**Questions :**

1. Construire l'arbre souhaité en utilisant l'algorithme ID3.
2. Reconstruire l'arbre en utilisant le Gini Index.
3. Donner le risque du patient ayant les attributs (Jeune,V,F) selon les deux arbres construits.

**Exercice 4** Considérons l'ensemble E d'exemples suivant ayant les attributs A,B,C et D :

	$E_1$	$E_2$	$E_3$	$E_4$	$E_5$	$E_6$	$E_7$	$E_8$	$E_9$	$E_{10}$
A	$a_1$	$a_1$	$a_1$	$a_2$	$a_2$	$a_2$	$a_1$	$a_2$	$a_3$	$a_3$
B	$b_1$	$b_2$	$b_2$	$b_1$	$b_2$	$b_2$	$b_1$	$b_1$	$b_1$	$b_2$
C	$c_1$	$c_2$	$c_3$	$c_1$	$c_1$	$c_1$	$c_1$	$c_2$	$c_3$	$c_2$
D	$d_2$	$d_2$	$d_1$	$d_1$	$d_1$	$d_2$	$d_1$	$d_2$	$d_1$	$d_2$
Classe	+	+	-	-	-	+	+	-	+	+

1. Construire l'arbre de décision correspondant à l'ensemble E en utilisant l'algorithme ID3.
2. Donner la précision de l'arbre obtenu sur la table d'entraînement. Calculer la moyenne harmonique sur la même table.

**NB :** On donne le tableau suivant représentant les valeurs de la fonctions

$$H(x, y) = -\frac{x}{x+y} \log_2\left(\frac{x}{x+y}\right) - \frac{y}{x+y} \log_2\left(\frac{y}{x+y}\right)$$

$x \setminus y$	1	2	3	4	5	6
6	0,592	0,811	0,918	0,971	0,994	1
5	0,650	0,863	0,954	0,991	1	
4	0,722	0,918	0,985	1		
3	0,811	0,971	1			
2	0,918	1				
1	1					

**Exercice 5** Nous considérons l'ensemble d'exemples représentant la nature de différents échantillons de champignons : toxique ou non selon les critères de couleur, taille, forme et le milieu de croissance :

Coleur	Taille	Forme	Milieu	Toxique
marron	petit	plat	terre	oui
jaune	petit	sphère	terre	oui
marron	moyen	conique	bois	non
blanc	moyen	sphère	terre	non
blanc	grand	plat	terre	non

1. Construire l'arbre de décision correspondant à cet ensemble en utilisant l'algorithme ID3.
2. Reconstruire l'arbre en utilisant le Gini Index.
3. Donner la précision des deux arbres et leurs moyennes harmoniques sur la table de test suivante :

Coleur	Taille	Forme	Milieu	Toxique
marron	grand	plat	bois	non
blanc	moyen	conique	terre	oui
jaune	moyen	sphère	terre	oui
marron	moyen	conique	bois	non
jaune	petit	plat	terre	oui

Responsable de la matière :  
**Dr A.Djeffal**