

## 0.1 Gestion d'attribut manquant

Exemple de gestion d'attributs à valeurs manquantes avec l'algorithme C4.5

Soit l'arbre de décision, ainsi que l'ensemble d'apprentissage suivants :

Temps	Température	Humidité	Vent	Jouer
Ensoleillé	Haute	Haute	Faux	Non
Ensoleillé	Haute	Haute	Vrai	Non
Couvert	Haute	Haute	Faux	Oui
Pluvieux	Basse	Haute	Faux	Oui
Pluvieux	Moyenne	Normal	Faux	Oui
Pluvieux	Moyenne	Normal	Vrai	Non
Couvert	Moyenne	Normal	Vrai	Oui
Ensoleillé	Basse	Haute	Faux	Non
Ensoleillé	Moyenne	Normal	Faux	Oui
Pluvieux	Basse	Normal	Faux	Oui
Ensoleillé	Basse	Normal	Vrai	Oui
Couvert	Basse	Haute	Vrai	Oui
Couvert	Haute	Normal	Faux	Oui
Pluvieux	Moyenne	Haute	Vrai	Non

TABLE 1 – Données d'apprentissage A

TABLE 2 – Sous ensemble F

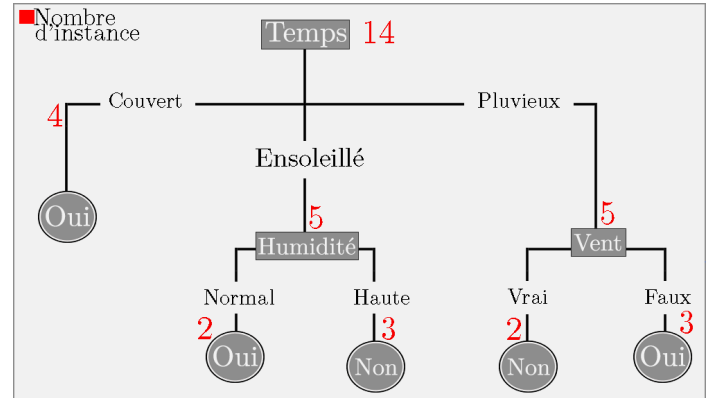


TABLE 3 – L'arbre de décision correspondant aux données A

Supposons que nous avons l'instance suivante à classer :

Temps	Température	Humidité	Vent	Jouer?
Pluvieux	Haute	Normal	.	?

Supposons qu'une instance des données de teste qui a comme valeur temps "Ensoleillé", mais qui n'a pas de valeur pour l'attribut humidité.

En parcourant les testes de notre arbre de décision, cette instance prendra le chemin "Pluvieux", ensuite, il y aura un teste pour lequel elle a une valeur manquante. Nous remarquons que le teste divise les données en deux sous ensemble, 2 instances corresponde au testes "vent=vrai", contre 3 instances qui correspondent au teste "vent=faux" L'algorithme C4.5 renverrait une distribution de probabilité de  $[\frac{2}{5}, \frac{3}{5}]$  correspondante à [Non,Oui] pour cette instance à valeur manquante.