

Compte-rendu du TP n°1

1) Le jeu de données comporte 14 instances qui sont composées de 5 attributs.

2) Nombre de valeurs distinctes par attribut :

Attribut	Nombre de valeurs
outlook	3
temperature	12
humidity	10
windy	2
play	2

3) Plage des valeurs des différents attributs numériques

Attribut numérique	Intervalle
temperature	[64, 85]
humidity	[65, 96]

4) La valeur d'attribut **overcast** permet de jouer dans 100 % des cas (play = yes).

5) Description de l'attribut play, en fonction de l'humidité.

Intervalle du taux d'humidité	Valeur la plus fréquente de play
[65, 80.5]	true
[80.5, 96]	false

6) Aucune autre valeur ne permet de donner à coup sûr, de déterminer la jouabilité ou non. On peut que donner des réponses probables mais pas certaines (i.e les histogrammes observés ne sont pas colorés soit en bleu, soit en rouge).

7) Comparaison des résultats obtenus lors des exécutions des différents algorithmes.

Algorithme	Attribut	Priorité
InfoGainAttributeEval	outlook	0.2467
	windy	0.0481
	humidity	0
	temperature	0
ReliefAttributeEval	outlook	0.05476
	humidity	0.01134
	windy	-0.00238
	temperature	-0.03137
GainRatioAttributeEval	outlook	0.1564
	windy	0.0488
	humidity	0
	temperature	0
SymetricalUnsertAttributeEval	outlook	0.196
	windy	0.05
	humidity	0
	temperature	0
OneRAttributeEval	humidity	57.143
	outlook	50
	temperature	50
	windy	42.857

8) A partir de ces données, on constate que l'ordre des attributs les plus fréquent est :

1. outlook
2. humidity
3. windy
4. temperature

9) Nombre de valeurs de l'attribut play, en fonction des couples de valeurs (outlook, windy).

outlook	windy	Play = yes	Play = no
sunny	TRUE	1	1
	FALSE	1	2
overcast	TRUE	2	0
	FALSE	2	0
rainy	TRUE	0	2
	FALSE	3	0

On peut alors déduire deux règles de prédiction supplémentaires : (rainy, TRUE) => (play = no) et (rainy, FALSE) => (play = yes).

10) Caractéristiques supplémentaires du point (sunny, FALSE), avec play = no :

Attribut	Valeur
temperature	80
humidity	90