

**TP1 Fouille de Données « Prise en main
de Weka »**



TP1 Fouille de Données

Guettouche Islem

Master 1 Génie de l'informatique logicielle (G.I.L)

Table des matières

Introduction :	3
Question1	3
Question 2	3
Question3	4
Question4	4
Question5	4
Question6	5
Question7	5
Question8	5
Question9	5
Question10	6

Introduction :

Le système Weka est développé à l'Université de Waikato, Nouvelle Zélande. Il permet de prétraiter des données, de les analyser à l'aide d'une méthode de data mining et d'afficher le modèle résultant et ses performances. Weka est entièrement développé en Java. Il est diffusé sous licence publique GNU. Les ressources nécessaires à l'installation et à l'utilisation du système sont disponibles à l'adresse suivante

<http://www.cs.waikato.ac.nz/ml/weka/>. Dans Weka, chaque méthode de transformation, de sélection d'attributs, d'apprentissage, de clustering ou de découverte d'associations est implémentée par une classe Java. La documentation des classes est accessible à partir du fichier packages.html. Weka traite des données au format ARFF (Attribute Relation Format File) ou CSV (Comma Separated Values).

Question1

- Le nombre d'attributs est : **5**.
- Le nombre d'instance est : **14**.

Question 2

- Nom de l'attribut: **outlook**.
- Type d'attribut: **nominal**.
- Nombre de valeurs manquantes: **0**.
- Nombre de valeurs distinctes: **3**.
- nombre de valeurs unique: **0**.
- Liste des valeurs : **sunny, overcast, rainy**.
- Nombre d'instances possédants chaque valeur:

 **Sunny : 5.**

 **Overcast: 4.**

 **Rainy : 5.**

- nom de l'attribut: **windy**.
- type d'attribut: **2**.
- Nombre de valeurs manquantes: **0**.
- Nombre de valeurs distinctes: **2**.
- nombre de valeurs unique: **0**.
- liste des valeurs : **true, false**.

la valeur de play la plus fréquente pour les instances Humidity $\in [80.5, 96]$. est : no

- Nombre d'instances possédants chaque valeur:

✚ True : 6.

✚ False : 8.

- Nom de l'attribut: **plasy**
- Type d'attribut: **nominal**
- Nombre de valeurs manquantes: **0**
- Nombre de valeurs distinctes: **2**
- Nombre de valeurs unique: **0**
- Liste des valeurs : **yes, no**
- Nombre d'instances possédants chaque valeur:

✚ Yes: 9

✚ No: 5

Question3

- Température :

✚ Minimal : 64

✚ Maximal : 85

- Humidité :

✚ Minimal : 65

✚ Maximal : 96

Question4

La valeur de l'attribut de outlook qui prédit qu'on peut tjrs joué est **overcast**.

Question5

La valeur de play la plus fréquente pour les instances Humidity [65, 80.5] est: **yes**

La valeur de play la plus fréquente pour les instances Humidity [80.5, 96] est : **no**

Question6

Aucun attribut associé à play qui pourra nous prédire à coup sûr que «**play**» sera à «**yes**»

Question7

PLAY

Algorithme	Ordre décroissant d'utilité
InfoGainAttributeEval	1,4,3,2
ReliefAttributeEval	1,3,4,2
GainRatioAttributeEval	1,4,3,2
SymmetricalUnserAttributeEval	1,4,3,2
OneRAttributeEval	3,1,2,4

Question8

L'ordre des attributs les plus fréquents est:

- Outlook
- Windy
- Humidité
- Température

Question9

- ✚ TRUE sunny: 2
- ✚ TRUE overcast: 2
- ✚ TRUE rainy: 2
- ✚ FALSE sunny: 3
- ✚ FALSE overcast: 2
- ✚ FALSE rainy: 3

Question10

- Instance: **2**
- Outlook : **sunny**
- Température : **80.0**
- Humidité : **90.0**
- Windy : **TRUE**
- Play : **no**