

# TP: Weka RA - BI

Durand JF & Amara Antoine

---

## Exercice 1

L'algorithme apriori nous retourne par défaut les 10 meilleures règles trouvées. Ci-dessous les règles retournées par weka.

Best rules found:

```
1. outlook=overcast 4 ==> play=yes 4    <conf:(1)> lift:(1.56) lev:(0.1) [1]
conv:(1.43)
2. temperature=cool 4 ==> humidity=normal 4    <conf:(1)> lift:(2) lev:(0.14) [2]
conv:(2)
3. humidity=normal windy=FALSE 4 ==> play=yes 4    <conf:(1)> lift:(1.56) lev:(0.1)
[1] conv:(1.43)
4. outlook=sunny play=no 3 ==> humidity=high 3    <conf:(1)> lift:(2) lev:(0.11)
[1] conv:(1.5)
5. outlook=sunny humidity=high 3 ==> play=no 3    <conf:(1)> lift:(2.8) lev:(0.14)
[1] conv:(1.93)
6. outlook=rainy play=yes 3 ==> windy=FALSE 3    <conf:(1)> lift:(1.75) lev:(0.09)
[1] conv:(1.29)
7. outlook=rainy windy=FALSE 3 ==> play=yes 3    <conf:(1)> lift:(1.56) lev:(0.08)
[1] conv:(1.07)
8. temperature=cool play=yes 3 ==> humidity=normal 3    <conf:(1)> lift:(2)
lev:(0.11) [1] conv:(1.5)
9. outlook=sunny temperature=hot 2 ==> humidity=high 2    <conf:(1)> lift:(2)
lev:(0.07) [1] conv:(1)
10. temperature=hot play=no 2 ==> outlook=sunny 2    <conf:(1)> lift:(2.8)
lev:(0.09) [1] conv:(1.29)
```

## Exercice 2

```
Calcul de la confiance = sup(R1) / sup(antécédent de R1)
                      = sup (p1->p2) / sup(p1)
                      = 4/4 = 1
```

```
Calcul de l'amélioration = Conf(AB) / p(B)
                      = 1 / (7/14)
                      = 1 / 0.5 = 2
```

```
Calcul du leverage = influence
```

```
Calcul de la conviction = influence
```

## Exercice 3

**a**

On ne peut pas lancer l'algorithme, il semblerait que ça soit causé par une spécificité du csv qui pourrait éventuellement fournir des lignes de tailles différentes et donc fausser le calcul de l'algorithme apriori.

**b**

La suppression de l'attribut id nous permet d'obtenir une visualisation.

**c**

```
@attribute children
{'\'(-inf-0.0]\'',\'\'(0.0-1.0]\'',\'\'(1.0-2.0]\'',\'\'(2.0-3.0]\'',\'\'(3.0-4.0]\'',\'\'(4.0-inf)\''}
```

**d**

Problème avec le fichier bank.arff apparemment ça ne serait pas la bonne manière de coder l'intervall de children.