UMB M'sila
Département de l'Informatique

Master2 SIGL
Module DMRI

TP N°5

Règles d'association avec R

Objectif du TP :

Appliquer la méthode Apriori pour la génération des règles d'association à partir d'un ensemble de données.

Etapes du TP

1- **Ensemble de données (le fichier Titanic) :** voyageurs du titanic par class, Sex, Age, Survived (Yes/No) et la fréquence (Freq) qui donne le nombre de voyageurs.

Class	Sex	Age	Survived	Freq
2nd	Male	Child	No	0
3rd	Male	Child	No	35
Crew	Male	Child	No	0
1st	Female	Child	No	0
2nd	Female	Child	No	0
3rd	Female	Child	No	17
Crew	Female	Child	No	0
1st	Male	Adult	No	118
2nd	Male	Adult	No	154
3rd	Male	Adult	No	387
Crew	Male	Adult	No	670
1st	Female	Adult	No	4
2nd	Female	Adult	No	13
3rd	Female	Adult	No	89

2- Packages à installer:

Nous allons utiliser le package suivant:

1. **arule** – Règles d'association

3- Pour que les données soit utilisables par l'algorithme d'extraction des règles d'associations, on doit les transformer de telle manière que chaque ligne représente une personne. La nouvelle structure des données est *titanic.raw*

```
str(Titanic)
df <- as.data.frame(Titanic)
```

```
head(df)
titanic.raw <- NULL
for(i in 1:4) {
  titanic.raw <- cbind(titanic.raw, rep(as.character(df[,i]), df$Freq))
}
titanic.raw <- as.data.frame(titanic.raw)
names(titanic.raw) <- names(df)[1:4]
dim(titanic.raw)
str(titanic.raw)
head(titanic.raw)
summary(titanic.raw)
```

4- Règles d'association : On applique l'algorithme APRIORI en appelant la fonction apriori() avec les paramètres par défaut suivant : *sup=0.1* (10%), *conf=0.8* (80%) et *maxlen=10* (Longueur maximale d'une règle). On doit trouver 27 règles.

```
library(arules)
# trouver les règles d'association en utilisant les paramètres par défaut
rules.all <- apriori(titanic.raw)
rules.all
inspect(rules.all)
```

```
lhs                                    rhs           support confidence coverage
[1]  {}                                => {Age=Adult}   0.950   0.950      1.000
[2]  {Class=2nd}                       => {Age=Adult}   0.119   0.916      0.129
[3]  {Class=1st}                       => {Age=Adult}   0.145   0.982      0.148
[4]  {Sex=Female}                      => {Age=Adult}   0.193   0.904      0.214
[5]  {Class=3rd}                       => {Age=Adult}   0.285   0.888      0.321
[6]  {Survived=Yes}                    => {Age=Adult}   0.297   0.920      0.323
[7]  {Class=Crew}                      => {Sex=Male}    0.392   0.974      0.402
[8]  {Class=Crew}                      => {Age=Adult}   0.402   1.000      0.402
[9]  {Survived=No}                     => {Sex=Male}    0.620   0.915      0.677
[10] {Survived=No}                     => {Age=Adult}   0.653   0.965      0.677
[11] {Sex=Male}                        => {Age=Adult}   0.757   0.963      0.786
[12] {Sex=Female,Survived=Yes}         => {Age=Adult}   0.144   0.919      0.156
[13] {Class=3rd,Sex=Male}              => {Survived=No} 0.192   0.827      0.232
[14] {Class=3rd,Survived=No}           => {Age=Adult}   0.216   0.902      0.240
[15] {Class=3rd,Sex=Male}              => {Age=Adult}   0.210   0.906      0.232
[16] {Sex=Male,Survived=Yes}           => {Age=Adult}   0.154   0.921      0.167
[17] {Class=Crew,Survived=No}          => {Sex=Male}    0.304   0.996      0.306
[18] {Class=Crew,Survived=No}          => {Age=Adult}   0.306   1.000      0.306
[19] {Class=Crew,Sex=Male}             => {Age=Adult}   0.392   1.000      0.392
[20] {Class=Crew,Age=Adult}            => {Sex=Male}    0.392   0.974      0.402
[21] {Sex=Male,Survived=No}            => {Age=Adult}   0.604   0.974      0.620
[22] {Age=Adult,Survived=No}           => {Sex=Male}    0.604   0.924      0.653
[23] {Class=3rd,Sex=Male,Survived=No}  => {Age=Adult}   0.176   0.917      0.192
[24] {Class=3rd,Age=Adult,Survived=No} => {Sex=Male}    0.176   0.813      0.216
[25] {Class=3rd,Sex=Male,Age=Adult}    => {Survived=No} 0.176   0.838      0.210
[26] {Class=Crew,Sex=Male,Survived=No} => {Age=Adult}   0.304   1.000      0.304
[27] {Class=Crew,Age=Adult,Survived=No} => {Sex=Male}   0.304   0.996      0.306
```

5- **Filtrage des règles :** On veut garder les règles dont la partie droite indique si la personne a survécue ou non. On écrit alors *rhs=c("Survived=No", "Survived=Yes")* dans le paramètre *appearance*. Remarquer qu'on a modifié la valeur du support. On trouvera 12 règles.

```
rules <- apriori(titanic.raw, control = list(verbose=F),
                 parameter = list(minlen=2, supp=0.005, conf=0.8),
                 appearance = list(rhs=c("Survived=No", "Survived=Yes"),
                                   default="lhs"))
quality(rules) <- round(quality(rules), digits=3)
rules.sorted <- sort(rules, by="lift")
inspect(rules.sorted)
```

```
lhs                                 rhs            support confidence coverage
[1]  {Class=2nd,Age=Child}             => {Survived=Yes} 0.011   1.000      0.011
[2]  {Class=2nd,Sex=Female,Age=Child}  => {Survived=Yes} 0.006   1.000      0.006
[3]  {Class=1st,Sex=Female}            => {Survived=Yes} 0.064   0.972      0.066
[4]  {Class=1st,Sex=Female,Age=Adult}  => {Survived=Yes} 0.064   0.972      0.065
[5]  {Class=2nd,Sex=Female}            => {Survived=Yes} 0.042   0.877      0.048
[6]  {Class=Crew,Sex=Female}           => {Survived=Yes} 0.009   0.870      0.010
[7]  {Class=Crew,Sex=Female,Age=Adult} => {Survived=Yes} 0.009   0.870      0.010
[8]  {Class=2nd,Sex=Female,Age=Adult}  => {Survived=Yes} 0.036   0.860      0.042
[9]  {Class=2nd,Sex=Male,Age=Adult}    => {Survived=No}  0.070   0.917      0.076
[10] {Class=2nd,Sex=Male}              => {Survived=No}  0.070   0.860      0.081
[11] {Class=3rd,Sex=Male,Age=Adult}    => {Survived=No}  0.176   0.838      0.210
[12] {Class=3rd,Sex=Male}              => {Survived=No}  0.192   0.827      0.232
```

Dr. Tahar MEHENNI