

LAB8

Carles Garriga Estrad & Balbina Virgili Rocosa

6/4/2018

Exercise 1

The tic_tt dataset has been read, by specifying the first column (identifying the rowId) as NULL and all the other columns as factors because all of them are categorical values.

```
tic_tt <- read.csv("tic_tt.csv", header = TRUE, sep = ";", stringsAsFactors = TRUE,
                  colClasses=c('NULL', rep('factor', 33)), fileEncoding="latin1")
```

Exercise 2

Having read the dataset, the catdesc function has been called using *Pagament.a.trav.s.d.Internet* as response variable.

```
tic_tt.catdes = catdes(as.data.frame(tic_tt), num.var = 28)
```

We can observe with the results retrieved, the variable “Ha.comprat.per.Internet” is the most highly correlated one to the response variable, because (usually) in order to buy a product on the internet you have to pay for that product. We can realize that it is overexpressed when “Ha.comprat.per.Internet.=TRUE” and very underexpressed when “Ha.comprat.per.Internet.=FALSE”. Furthermore, the response variable is correlated with services from banks (such as transfers or payments) which do not necessarily involve a product. After this variable “Amb.quina.freq.ncia.usa.inet.=Diariament” is the next important one, defining that if an individual uses internet everyday, could lead him / her to make payments online. To sum up, we can say that people who do payments by internet are the ones that daily use internet, they have bought products using internet, such as food, but also they use electronic bank services.

Exercise 3

In order to convert the dataset into a transactions file, we have used the as method as follows.

```
ttr <- as(tic_tt, "transactions")
```

Exercise 4

Before running the apriori algorithm in order to compute the association rules, we have to define several parameters that are going to be used. We have defined the minimum support to 0.1 as well as the minimum confidence to 0.7. Finally, we have settled the maximum items to be 15. Finally the association rules are computed using the apriori algorithm.

Exercise 5

Using the previous association rules, the frequent itemsets are computed using the generatingItemsets method. The ten most significant frequent Itemsets are displayed ordered using the support (decreasing).

items **support**

- [1] {Tr.mits.amb.els.serveis.sanitaris.=FALSE} **0.9784946**
- [2] {Fa.teletreball.=FALSE} **0.9668459**
- [3] {Ha.comprat.aliments.=FALSE} **0.9623656**
- [4] {Fa.teletreball.=FALSE,Tr.mits.amb.els.serveis.sanitaris.=FALSE} **0.9453405**
- [5] {Ha.comprat.aliments.=FALSE,Tr.mits.amb.els.serveis.sanitaris.=FALSE} **0.9426523**

- [6] {Fa.teletreball.=FALSE,Ha.comprat.aliments.=FALSE} **0.9327957**
- [7] {Estudia.per.Internet.=FALSE} **0.9301075**
- [8] {Estudia.per.Internet.=FALSE,Tr.mits.amb.els.serveis.sanitaris.=FALSE} **0.9130824**
- [9] {Fa.teletreball.=FALSE,Ha.comprat.aliments.=FALSE,Tr.mits.amb.els.serveis.sanitaris.=FALSE} **0.9130824**
- [10] {Estudia.per.Internet.=FALSE,Fa.teletreball.=FALSE} **0.9050179**

Exercise 6

Also, the ten most significant association rules ordered by the lift (decreasing) are shown.

lhs rhs support confidence lift count

- [1] {Disposa.de.connexi..a.Internet.a.la.llar.=TRUE,
Fa.servir.correu.electr.nic.=TRUE,
Ha.comprat.per.Internet.=TRUE,
Visita.webs.de.l.Adm..=TRUE} => {Pagament.a.trav.s.d.Internet.=TRUE}
- 0.1021505 0.7549669 5.506817 114
- [2] {Disposa.de.connexi..a.Internet.a.la.llar.=TRUE,
Ha.comprat.per.Internet.=TRUE,
Tr.mits.amb.els.serveis.sanitaris.=FALSE,
Visita.webs.de.l.Adm..=TRUE} => {Pagament.a.trav.s.d.Internet.=TRUE}
- 0.1012545 0.7533333 5.494902 113
- [3] {Disposa.de.connexi..a.Internet.a.la.llar.=TRUE,
Ha.comprat.per.Internet.=TRUE,
Visita.webs.de.l.Adm..=TRUE} => {Pagament.a.trav.s.d.Internet.=TRUE}
- 0.1066308 0.7531646 5.493671 119
- [4] {Disposa.de.connexi..a.Internet.a.la.llar.=TRUE,
Amb.quina.freq..ncia.usa.ordinador.=Diariament,
Ha.comprat.per.Internet.=TRUE} => {Pagament.a.trav.s.d.Internet.=TRUE}
- 0.1048387 0.7500000 5.470588 117
- [5] {Fa.servir.correu.electr.nic.=TRUE,
Ha.comprat.per.Internet.=TRUE,
Visita.webs.sanitaris.=FALSE} => {Pagament.a.trav.s.d.Internet.=TRUE}
- 0.1021505 0.7500000 5.470588 114
- [6] {Fa.servir.correu.electr.nic.=TRUE,
Ha.comprat.per.Internet.=TRUE,
Tr.mits.amb.els.serveis.sanitaris.=FALSE,
Visita.webs.sanitaris.=FALSE} => {Pagament.a.trav.s.d.Internet.=TRUE}
- 0.1021505 0.7500000 5.470588 114
- [7] {Disposa.de.connexi..a.Internet.a.la.llar.=TRUE,
Amb.quina.freq..ncia.usa.ordinador.=Diariament,
Fa.servir.correu.electr.nic.=TRUE,
Ha.comprat.per.Internet.=TRUE} => {Pagament.a.trav.s.d.Internet.=TRUE}
- 0.1012545 0.7483444 5.458512 113
- [8] {Fa.teletreball.=FALSE,
Ha.comprat.aliments.=FALSE,

Ha.comprat.per.Internet.=TRUE,
Tr.mits.amb.els.serveis.sanitaris.=FALSE} => {Pagament.a.trav.s.d.Internet.=TRUE}

0.1012545 0.7483444 5.458512 113

[9] {Fa.teletreball.=FALSE,
Ha.comprat.per.Internet.=TRUE,
Visita.webs.sanitaris.=FALSE} => {Pagament.a.trav.s.d.Internet.=TRUE}

0.1003584 0.7466667 5.446275 112

[10] {Fa.teletreball.=FALSE,
Ha.comprat.per.Internet.=TRUE,
Tr.mits.amb.els.serveis.sanitaris.=FALSE,
Visita.webs.sanitaris.=FALSE} => {Pagament.a.trav.s.d.Internet.=TRUE}

0.1003584 0.7466667 5.446275 112

Exercise 7

Finally, we want to filter the most significant association rules to ones that have as the right hand side the response value: “Pagament.a.trav.s.d.Internet”. Later on, those filtered rules are sorted by lift and displayed (showing only the ten first ones).

lhs rhs support confidence lift count

[1] {Disposa.de.connexi..a.Internet.a.la.llar.=TRUE,
Fa.servir.correu.electr.nic.=TRUE,
Ha.comprat.per.Internet.=TRUE,
Visita.webs.de.l.Adm..=TRUE} => {Pagament.a.trav.s.d.Internet.=TRUE}

0.1021505 0.7549669 5.506817 114

[2] {Disposa.de.connexi..a.Internet.a.la.llar.=TRUE,
Ha.comprat.per.Internet.=TRUE,
Tr.mits.amb.els.serveis.sanitaris.=FALSE,
Visita.webs.de.l.Adm..=TRUE} => {Pagament.a.trav.s.d.Internet.=TRUE}

0.1012545 0.7533333 5.494902 113

[3] {Disposa.de.connexi..a.Internet.a.la.llar.=TRUE,
Ha.comprat.per.Internet.=TRUE,
Visita.webs.de.l.Adm..=TRUE} => {Pagament.a.trav.s.d.Internet.=TRUE}

0.1066308 0.7531646 5.493671 119

[4] {Disposa.de.connexi..a.Internet.a.la.llar.=TRUE,
Amb.quina.freq.ncia.usa.ordinador.=Diariament,
Ha.comprat.per.Internet.=TRUE} => {Pagament.a.trav.s.d.Internet.=TRUE}

0.1048387 0.7500000 5.470588 117

[5] {Fa.servir.correu.electr.nic.=TRUE,
Ha.comprat.per.Internet.=TRUE,
Visita.webs.sanitaris.=FALSE} => {Pagament.a.trav.s.d.Internet.=TRUE}

0.1021505 0.7500000 5.470588 114

[6] {Fa.servir.correu.electr.nic.=TRUE,
Ha.comprat.per.Internet.=TRUE,
Tr.mits.amb.els.serveis.sanitaris.=FALSE,
Visita.webs.sanitaris.=FALSE} => {Pagament.a.trav.s.d.Internet.=TRUE}

0.1021505 0.7500000 5.470588 114

[7] {Disposa.de.connexi..a.Internet.a.la.llar.=TRUE,
 Amb.quina.freq..ncia.usa.ordinador.=Diariament,
 Fa.servir.correu.electr.nic.=TRUE,
 Ha.comprat.per.Internet.=TRUE} => {Pagament.a.trav.s.d.Internet.=TRUE}

0.1012545 0.7483444 5.458512 113

[8] {Fa.teletreball.=FALSE,
 Ha.comprat.aliments.=FALSE,
 Ha.comprat.per.Internet.=TRUE,
 Tr.mits.amb.els.serveis.sanitaris.=FALSE} => {Pagament.a.trav.s.d.Internet.=TRUE}

0.1012545 0.7483444 5.458512 113

[9] {Fa.teletreball.=FALSE,
 Ha.comprat.per.Internet.=TRUE,
 Visita.webs.sanitaris.=FALSE} => {Pagament.a.trav.s.d.Internet.=TRUE}

0.1003584 0.7466667 5.446275 112

[10] {Fa.teletreball.=FALSE,
 Ha.comprat.per.Internet.=TRUE,
 Tr.mits.amb.els.serveis.sanitaris.=FALSE,
 Visita.webs.sanitaris.=FALSE} => {Pagament.a.trav.s.d.Internet.=TRUE}

0.1003584 0.7466667 5.446275 112