Doctoral Program MAP-i
Advanced Topics in Data Mining and Logic Programming (2007/2008)
Association Rules Project

A site provides articles to its costumers. The site log data (which will be supplied by the lecturers) contains the registered access to the articles by its costumers in the site for one year. The administrator of the site needs to:

a) identify the sets of interrelated articles to propose a set of 5 thematic dossiers. These dossiers must hold some impact, that is must be read by a significant number of persons.

b) identify a set of interesting rules (roughly around 10 rules) that enable to increase the probability of a reader to consult low accessed articles.

c) present a short study on the use of different interest measures and its impact on the recommendation of new cases (new accesses).

d) write a report (max 4 pages) describing how the results were obtained.

1 Tasks (generic)

Describe the performed tasks to obtain the rules, namely:

- deriving the rules,

- parameters tuning,

- how rules were selected,

- rules interpretation

2 About the data

The supplied data must be use only for this project. It is not for free distribution. Thus, the data will not be available on line. To do this project send an email to the lecturers (pja@di.uminho.pt or amjorge@fep.up.pt).

3 Suggestions

Use the software *CAREN* available at www.di.uminho.pt/∼pja/class/caren.html