Aula8 – Regras de Associação (Subgrupos).

1. Considere o dataset adult (no blackboard). Gere regras para o atributo ***education*** e indique as características de cada tipo de formação académica.

Resposta

**>** *caren adult.data 0.01 0.5 -s, -ovrt -Att -Heducation -Fisher*

Aplicar a discretização de fayyad Irani aos dados contínuos age.

**>** *caren adult.data 0.01 0.5 -s, -ovrt -Att -Heducation -fisher*

*-classeducation -Discfiage,education.num*

**>** *caren adult.data 0.01 0.5 -s, -ovrexe2.3 -Att -fisher -classeducation -Srikage,education.num*

1. Gere regras “contrast sets” para 3 grupos definidos pelo atributo *education* (*Masters*, *Bachelors* e *Doctorate*). Compare os resultados com as regras derivadas na alínea anterior.

*java caren adult.data 0.01 0.5 -s, -ovrt -Att -heducation=Masters,education=Bachelors,education=Doctorate -fisher -classeducation -Discfiage,education.num*

Com contraste “-CS”

*java caren adult.data 0.01 0.5 -s, -ovrexe2.4 -Att -heducation=Masters,education=Bachelors,education=Doctorate -classeducation -Discfiage -CS*

(Olhando para o atributo sex)

*> caren adult.data 0.01 0.5 -s, -ovrexe2.5 -Att -hsex=Male,sex=Female -classsex -Discfiage -CS*

Discretização idade e anos de educação

> caren adult.data 0.01 0.5 -s, -ovrexe2.6 -Att -hsex=Male,sex=Female -classsex -Discfiage,education.num -CS

1. Indique as características que ajudam a explicar o contraste entre os dois grupos de rendimento (i.e. *>50K* e *<=50K*).

Resposta

*> caren adult.data 0.01 0.5 -s, -ovrexe2.7 -Att -hclass=>50K,class=<=50K -classclass -fisher -Discfiage -CS*

1. Descarregue o dataset wage85 (ver blackboard – o ficheiro *wagesmicrodata.xls* contém explicações sobre os valores no dataset). Considere a propriedade numérica *wage* (rendimento/salário por hora). Apresente regras interessantes sobre rendimento dos mais jovens e das pessoas com mais formação académica. Apresente também casos de discriminação clara (minorias)!

Resposta

*java caren wages85.csv 0.01 0.5 -s; -ovrexe4 -Att -GWAGE -imp0.000000001 -classWAGE -Discbuk3 -DiscfiAGE,EXPERIENCE,EDUCATION -classWAGE*