

Podniková analytika (Cvičenie – deskriptívne DM)

Úlohy na zhukovanie

1. K-means (Mtcars)

- Načítajte dáta **mtcars** do premennej s názvom **mtcars1** (dáta mtcars nájdete napr. aj tu <http://peter.butka.website.tuke.sk/res/mtcars.csv>)
- Vyberte z dát **mtcars1** len stĺpce s názvom **hp** a **drat**
- Pomocou algoritmu k-means rozdeľte dáta **mtcars1** do dvoch zhukov (tried) a uložte tieto výsledky do premennej s názvom **kmeans.results1**, zobrazte aj prvky modelu (napr. centroidy, priradenia k zhukom, a pod.)
- Vytvorte kontingenčnú tabuľku (obdobu confusion matrix), ktorá bude porovnávať hodnoty atribútu **vs** z dát **mtcars** a získané klastre (zhuky) uložené v premennej **kmeans.results1**
- Vytvorte graf, v ktorom budú znázornené atribúty **hp** a **drat** a vykreslené body rozdeľte farebne podľa získaných zhukov
- Doplníte do grafu body centroidov pre jednotlivé zhuky

2. Hierarchické zhukovanie (Mtcars)

- Uložte do premennej s názvom **mtcars2** všetky riadky dát **mtcars** a stĺpce **mpg** a **qsec**
- Do premennej s názvom **hc2** uložte výsledky hierarchického zhukovania z dát **mtcars2**, metódu tohto zhukovania nastavte tak, že algoritmus bude brať vzdialenosť medzi zhukmi ako *priemer* vzdialenosti bodov v jednom zhuku a bodov v inom zhuku, použitie aglomeratívne zhukovanie
- Vykreslite hierarchické prepájanie zhukov (klastrov) pomocou dendrogramu
- Orežte vykreslený dendrogram na úroveň troch zhukov
- Pridajte do dát **mtcars** nový stĺpec s názvom **groups**, ktorý bude obsahovať ID získaných zhukov

Úloha na vytvorenie asociačných pravidiel

3. Asociačné pravidlá (AdultUCI)

- Načítajte dáta **AdultUCI** (dáta nájdete tu: http://peter.butka.website.tuke.sk/res/AdultUCI_world.csv)
- Z dát **AdultUCI** vytvorte jej podmnožinu s názvom **AdultUCI1**, ktorá bude obsahovať len stĺpce dátového typu faktor (samozrejme skontrolujte všetky typy a podľa potreby transformujte ak niektoré boli zle načítané, ale je jednoduché ich opraviť na správny typ)
- Vytvorte premennú **rules1**, ktorá bude obsahovať asociačné pravidlá získané z týchto dát pomocou algoritmu *apriori*, minimálnu dĺžku pravidla nastavte na 4, minimálnu podporu pravidiel na 0,01, minimálnu spoľahlivosť na 0,68, sledovaný atribút (pravá strana pravidla) bude obsahovať možnosti atribútu **income**
- Získané pravidlá usporiadajte podľa ukazovateľa **Lift** a následne ich vypíšte
- Vymažte zo získaných pravidiel tie, ktoré sú redundantné a na záver vypíšte opäť všetky už neredundantné pravidlá