

Analiza tabelaričnih podatkov

Navodila za izdelavo seminarske naloge

1. Vsak študent si za izbrane tabelarične podatke¹ zastavi nalogo odkrivanja zakonitosti v podatkih. Temi dveh študentov sta si lahko podobni, ne pa enaki. Temo prijavi Petri Kralj.
2. Podatke pripravi v primeren format za analizo s programom Weka².
3. Analiza podatkov z vsaj eno metodo za analizo tabelaričnih podatkov, ki smo jo uporabili na vajah.
4. V poročilu naj bo opisan problem, podatki, metode in rezultati. Poročilo naj bo v formatu prispevkov za konferenco Informacijska družba.
http://is.ijs.si/main_slo.html --> Primer članka
Članek je lahko v slovenskem ali angleškem jeziku. Držite se predpisanega oblikovanja in omejitve na 4 strani.

Predlog strukture članka:

- Povzetek: do 200 besed povzetka celotnega članka
- Uvod: vir in pomembnost zastavljenega problema, načrt postopka rešitve
- Podatki: vir podatkov, opis podatkov, opis potrebnega predprocesiranja, število primerov in atributov, opis ciljne spremenljivke (če je v podatkih) in njene distribucije, število manjkajočih vrednosti, ...
- Metode: kratek opis uporabljenih metod in parametri
- Rezultati - odvisno od izbrane metode: vizualizacija modela ali izpis pravil, ocena klasifikacijske točnosti ali podpore in zaupanja. Če je mogoče tudi interpretacija. So rezultati dobri?
- Zaključek: Kaj novega, zanimivega, uporabnega, koristnega ste odkrili, kaj bi se še dalo narediti.
- Viri: če boste uporabljali Weko

Witten, I. (et al.): Data Mining: Practical machine learning tools and techniques with Java implementations. Second Edition, 2005, Morgan Kaufman

Poročilo oddate **do vključno 7.11.2007** po elektronski pošti na naslov Petra.Kralj@ijs.si.

5. Predstavitev seminarske naloge iz analize tabelaričnih podatkov bo 11.5.2007. Vsak študent pripravi 5 minutno³ predstavitev v *MS PowerPoint* z uporabo *.pot* šablone. Poudarek naj bo na vaših podatkih in vaših rešitvah, ne pa na metodah. S seboj prinese tiskano in elektronsko verzijo poročila. Bodite pripravljene odgovarjati na vprašanja profesorjev in študentov.

Ocenjujemo: kvaliteto opravljenega dela, kvaliteto pisne predstavitve, kvaliteto ustne predstavitve in odgovore na vprašanja.

Za kakršnokoli vprašanje sem vam na voljo na elektronskem naslovu Petra.Kralj@ijs.si.

Gradivo, napotki in ideje so dostopni na spletni strani

<http://kt.ijs.si/PetraKralj/UNGKnowledgeDiscovery0607.html>

¹ Tabelarični podatki naj imajo nekaj deset atributov (stolpcev) in nekaj sto primerov (vrstic). Ciljna spremenljivka je lahko nominalna (diskretna) ali zvezna.

² Po želji lahko analizirate tudi s kakšnim drugim programom za odkrivanje zakonitosti v podatkih.

³ Odstopanja od so nezaželena in bojo kaznovana.