

## Adatbányászat tételsor

2024

1. Az adatbányászat definíciói, alapfogalmai, példák. Az adatbányászat eredete. Alapfeladatok: felügyelt és nem-felügyelt adatbányászat. Alkalmazási területek. Kihívások. Adatfeltárás: leíró statisztikák és vizualizáció.  
(Slides: Bevezetés + Feltárás)
2. Adatbányászati folyamat. Adattípusok, mérési skálák. Adatállomány típusok. Adatminőségi problémák. Előfeldolgozás: aggregálás, mintavétel, dimenziócsökkentés, jellemző szelekció és transzformáció. OLAP eszközök.  
(Slides: Adatok + Feltárás)
3. Osztályozási alapfogalmak. Döntési fa alapú osztályozók. Fa építési algoritmusok: Hunt, CART, CHAID, C4.5. Osztályozók kiértékelése: tévesztési mátrix, metrikák, ROC görbe.  
(Slides: Osztályozás: Alapfogalmak, Döntési fák és Kiértékelés)
4. Osztályozási módszerek I: szabály alapú osztályozók, naív Bayes módszer,  $k$  legközelebbi szomszéd ( $k$ -NN).  
(Slide: Osztályozás: További módszerek)
5. Osztályozási módszerek: mesterséges neuron hálók (ANN), támaszvektor gépek (SVM), logisztikus regresszió, együttes osztályozók.  
(Slides: Osztályozás: További módszerek)
6. Hasonlóság és távolság. Klaszterezés: definíciók,  $K$ -közép módszer és variánsai. A klaszterezés kiértékelése: hasonlósági mátrix, korreláció, SSE, árnyék együtthető.  
(Slides: Adatok vége + Klaszterezés: Alapfogalmak és algoritmusok)
7. Klaszterezés: hierarchikus és sűrűség alapú módszerek. Egyszerű, teljes és átlagos kapcsolás, dendrogram, DBSCAN.  
(Slides: Klaszterezés: Alapfogalmak és algoritmusok)
8. Vásárlói kosár adatok. Gyakori tételcsoport és támogatottság. Az Apriori elv és algoritmus. Jelöltgenerálás. További gyakori tételcsoport kereső algoritmusok. Asszociációs szabályok. Megbízhatóság. Mintázat-kiértékelés: statisztikus mérőszámok, lift érték.  
(Slides: Asszociációs szabályok: Alapfogalmak és algoritmusok)
9. Rendellenesség-keresés: definíció, feladatok, alkalmazások. Módszerek: grafikus, statisztikus, távolság alapú, modell alapú. A téves riasztás problémája.  
(Slides: Rendellenesség keresés)

A levelező képzésen a 2. tétel OLAP eszközeiből és a 9. tételből nem kell felkészülni.

Debrecen, 2024. május 20.

Dr. Ispány Márton