

**Komplex vizsga főtárgy**      Információ technológia, Adattudomány

**Tematika**

Nagy mennyiségű adat hatékony kezelése.

Adatbányászat modellek: regressziós modellek és regularizáció, kernel módszer és radiális bázisfüggvények, ritka kernelek (SVM és RVM), grafikus modellek és Bayes-hálók, nagy-dimenziós feladatok. Ezeken belül külön figyelemmel a legmodernebb sztochasztikus optimalizálókra, mint pl. a sztochasztikus gradiens típusú módszerek és Bayesi elven működő eljárások, pl. nemparaméteres tanulás. Nem-felügyelt adatbányászati modellek: keverékek és az EM algoritmus, klaszterezés, Kohonen-féle hálók, főkomponens-analízis és variánsai (pl. kernel-PCA), szinguláris felbontás, nem-negatív mátrix faktorizáció, független komponens analízis, többdimenziós skálázás. Egy adattudományi rendszer, pl. az Anaconda Python disztribúciójának használata.

Adatbányászat; tudásfeltárás adatbázisokban; szimbolikus adatbányászati módszerek. Gyakori minták; gyakori asszociációs szabályok. Gyakori mintákat kereső algoritmusok: Apriori, Apriori-Close, Eclat, Charm, Touch. Galois-hálók, Galois-hálókat előállító algoritmusok. A Snow algoritmus; hipergráfok. Ritka minták, ritka asszociációs szabályok. Szintenkénti és mélységi stratégiát alkalmazó algoritmusok ritka minták keresésére: Apriori-Rare, Walky-G. Esettanulmányok; a Coron rendszer

**Irodalom**

1. Bishop, C. M., Pattern Recognition and Machine Learning, Springer, 2006.
2. Hastie, T., Tibshirani, R., Friedman, J., The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer-Verlag, 2009.
3. Koski, T., Noble, J.M., Bayesian Networks. An Introduction. Wiley, 2009.
4. Gorelick, Micha, Ozsvald, Ian, High Performance Python: Practical Performant Programming for Humans (1st ed.). O'Reilly Media, 2014
5. Tan, P.-N., Steinbach, M., Karpatne, A., Kumar, V.: Introduction to Data Mining, 2nd ed., Pearson, 2018.
6. Han, J., Kamber, M., Pei, J.: Data Mining: Concepts and Techniques, 3rd ed., Morgan Kaufmann, 2011.
7. Liu, B.: Web Data Mining, Exploring Hyperlinks, Contents, and Usage Data, 2nd ed., Springer 2011.

**A főtárgy választásához  
kötelezően elvégzendő  
tantárgyak**

A programvezető jóváhagyásával:

1) Az alábbi tárgyak közül választandó négy tárgy:

- Internetes alkalmazások modern megoldásai (Adamkó Attila)
- Elosztott információtechnológiai rendszerek statisztikai analízise (Gál Zoltán)
- Haladó adatbányászat és alkalmazásai (Ispány Márton)
- Sztochasztikus adatbányászat és alkalmazásai (Ispány Márton)
- Szimbolikus adatbányászat (Szathmáry László)
- Statisztika alkalmazása az információ technológia területén (Terdik György)
- Statisztika, idősorok és alkalmazásai az informatikában (Terdik György)

2) Az Informatikai Doktori Iskola más programjaiból választandó két tárgy.

3) A magyarországi Doktori Iskolák programjaiból választandó egy tárgy.

**A főtárgy választásához  
ajánlottan elvégzendő  
tantárgyak**