

Óbudai Egyetem Neumann János Informatikai Kar		Kiberfizikai rendszerek intézet		
Tantárgy neve és kódja: Bevezetés az adattudományba / Mérnök Informatikus BSc szak		Kreditérték: 5 Nappali tagozat 2021/22 tanév I. félév		
Tantárgy oktató: Kail Eszter, Fleiner Rita, Piros Péter				
Előtanulmányi feltételek: (kóddal)		Adatbázisok NIXAB0HBNE		
Heti óraszámok:	Előadás: 2	Tantermi gyak.: 0	Laborgyakorlat: 2	Konzultáció: 0
Számonkérés módja:	Vizsga			
A tananyag				
Oktatási cél: A tárgy célja az adattudomány alapfogalmainak és folyamatainak gyakorlati megközelítésű megismertetése. A hallgatók a gyakorlati életből vett valós alkalmazási példákon keresztül az ismereteket megtapasztalva, egyre mélyebbre haladva precíz elméleti és egyúttal praktikus gyakorlati ismeretekhez jutnak. Az elméleti ismeretek gerincét a gépi tanulás algoritmusai adják, a gyakorlati feladatok pedig az elmélet gyakorlati megjelenítését szolgálják Python nyelv használatán keresztül.				
Tematika: Előadás: Történet, példák, esettanulmányok, az adattudományba sorolható területek. CRISP DM és a 6 fázisa. Adattisztítás. Felügyelt, felügyelet nélküli és megerősített tanulás. Túltanulás, alultanulás. Modellek validálása, tanulási/validációs/tesztelési halmaz, cross-validáció, Bias-Variance, ROC görbe. Legkisebb négyzetek módszere. Lineáris Regresszió. Gradiens módszer, maximum-likelihood becslés. Logisztikus regresszió, Tanulási/validációs/tesztelési halmaz, cross-validáció, Bias-Variance tradeoff, Precision-Recall, F1-score, ROC görbe. SVM. Neurális hálók. Döntési fák. Véletlen erdők. Boosting. Nem Ellenőrzött tanulás. Klaszterezés. K-means klaszterezés. Megerősített tanulás. Nagyobb esettanulmányok, kitekintés. Gyakorlat: Az adatmanipulálás, prediktív analízis, megjelenítés lépései valódi adatokkal elsősorban Python-csomagok (pandas, scikit-learn, matplotlib, ggplot) használatával.				

Féléves ütemezés:	
Oktatási hét (konzultáció)	Témakör
1.	Történet, példák, alapfogalmak, esettanulmányok
2.	Statisztika (gyakoriságok, percentilisek, átlag, medián, szórás, kovariancia, korreláció, diagramtípusok)
3.	Adatbányászat folyamata – CRISP DM
4.	Felügyelt tanulás, modellek kiértékelése – Tanító-, teszt- és validációs halmaz
5.	Kereszt validáció
6.	Lineáris és logisztikus regresszió
7.	Döntési fák
8.	Bagging, Random Forest, Boosting
9.	SVM, Neurális hálók
10.	Nem ellenőrzött tanulás, klasztizálás
11.	Megerősített tanulás
12.	Megerősített tanulás
13.	ZH írás
14.	Pót ZH írás
Félévközi követelmények	
A félév során a hallgatóka 13. héten zárthelyi dolgozatot írnak az előadás és a labor anyagából. A dolgozatokon 50-50 pontot lehet szerezni, ezek összegéből áll össze a végső pontszám (max. 100 pont). Mindkét ZH-n minimum 26 pontot kell teljesíteni a kurzus sikeres elvégzéséhez.	
Zárthelyi dolgozatok	
Oktatási hét (konzultáció)	Témakör

13	Előadás és labor ZH
14	ZH-k pótlása
1	Szöveg beírásához kattintson ide.
A félévzáró érdemjegy kialakításának módszere	
A félévzáró érdemjegy ZH-k pontszáma alapján áll össze, amikre maximum 100 pont szerezhető. Az elégséges jegyhez 52, a közepeshez 63, a jóhoz 74, a jeleshez 85 pontot kell elérni.	
Pótlás módja	
Az utolsó gyakorlaton az egyik ZH pótolható. Évközi jegy pótláson mindkét ZH pótolható. Itt is érvényes, hogy minimum 26 pontot kell teljesíteni mindkét ZH-n.)	
Vizsga módja	
Szöveg beírásához kattintson ide.	
Vizsgajegy kialakítása	
Az egyes érdemjegyek ponthatára	
Irodalom	
Kötelező:	
A Moodle rendszerben közzétett jegyzet	
Ajánlott:	
Richard S. Sutton and Andrew G. Barto: Reinforcement Learning: An Introduction	
J. Leskovec, A. Rajaraman, J. D. Ullman: Mining of Massive Datasets	
J. Gareth et al.: An Introduction to Statistical Learning	
Szöveg beírásához kattintson ide.	
Egyéb segédletek:	