

Óbudai Egyetem Neumann János Informatikai Kar		Kiberfizikai Rendszerek Intézet		
Tantárgy neve és kódja: Bevezetés az adattudományba NKXBATHBNE Kreditérték: 5 / Mérnök Informatikus BSc szakNappali tagozat 2022/23 tanév I. félév				
Tantárgy felelős: Balázsné Dr. Kail Eszter Tantárgy oktató: Balázsné Dr. Kail Eszter, Piros Péter				
Előtanulmányi feltételek: (kóddal)		NIXSS1HBNE Szakmai szigorlat		
Heti óraszámok:	Előadás: 2	Tantermi gyak.: 0	Laborgyakorlat: 2	Konzultáció: 0
Számonkérés módja:	Évközi jegy			
A tananyag				
Oktatási cél: 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: Történet, példák, esettanulmányok, az adattudományba sorolható területek. Statisztika alapok. CRISP DM és a 6 fázisa. Adattisztítás. Felügyelt, felügyelet nélküli és megerősítéses tanulás. Túltanulás, alultanulás. Modellek validálása, tanulási/validációs/tesztelési halmaz, keresztvalidáció, Bias-Variance, Precision-Recall, F1-score ROC görbe, AUC. Legkisebb négyzetek módszere. Lineáris Regresszió. Logisztikus regresszió. SVM. Neurális hálók. Döntési fák. Véletlen erdők. Boosting. Felügyelet nélküli tanulás. Klaszterezés. K-means klaszterezés. Megerősítéses tanulás. Nagyobb esettanulmányok, kitekintés. 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- Keresztvalidáció, bootstrap
5.	Lineáris és logisztikus regresszió
6.	Döntési fák, Bagging, Random Forest, Boosting
7.	SVM
8.	Neurális hálók
9.	Nem felügyelt tanulás, klaszterezés I.
10.	Nem felügyelt tanulás, klaszterezés II.
11.	Megerősítéses tanulás I.
12.	Megerősítéses tanulás II.
13.	Elméleti ZH, labor ZH
14.	Pót ZH írás
Félévközi követelmények	
A félév során a hallgatók a 13. héten egy elméleti és egy labor 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	ZH (elméleti + labor)
14	Pót ZH
A félévzáró érdemjegy kialakításának módszere	
A félévzáró érdemjegy a ZH-k pontszáma alapján áll össze, amikre maximum 100 pont kapható. 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	
-	
Vizsgajegy kialakítása	
-	
Irodalom	
Kötelező:	
Han, J., & Kamber, M. (2004). Adatbányászat. Koncepciók és technikák. Panem, Budapest. 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	
Egyéb segédletek:	
Egyéb segédletek felsorolása, elérhetőségük megnevezése (pl.: http://nik.uni-obuda.hu/oop)	