

Kiberfizikai Rendszerek Intézet			Választható 25-26-2			
Tantárgy neve:	Kódja:	Kredit:	Óraszám			
				ea	tg	lab
Bevezetés az adattudományba	NKWBA1HBLF	5	levelező, féléves	10	0	10
Tárgyfelelős: Dr. habil. Dineva Adrienn Alexandra			Beosztás: egyetemi adjunktus			
Oktató(k): Angyalné Dr. Alexy Márta, Dr. Alwahab Dhulfiqar Zoltán						
Előtanulmányi feltételek:	NIXAB0HBLE	Adatbázisok				
Számonkérés módja:	évközi jegy					
A tananyag						
Oktatási cél:	A tárgy célja az adattudomány alapfogalmainak, alapvető módszereinek és folyamatainak gyakorlati megközelítésű megismertetése. A hallgatók a gyakorlati életből vett valós alkalmazási példákon keresztül az ismereteket megtapasztalva, elméleti és egyúttal praktikus gyakorlati ismeretekhez jutnak. Az elméleti ismeretek gerincét a felügyelt gépi tanuló algoritmusok adják, a gyakorlati feladatok pedig az elmélet gyakorlati megjelenítését szolgálják Python nyelv használatán keresztül.					
Tematika:	Előadás: Bevezetés, fogalmak, csoportosítás, klasszikus adatbányászati technikák. Általános elemző (adattudományi) folyamat. Adattisztítás, adatelőkészítés. Felügyelt és nem felügyelt gépi tanuló modellek. Modellek validálása. Labor: Az adatmanipulálás, prediktív analízis, megjelenítés lépései változatos adatokkal elsősorban Python-csomagok használatával.					

Féléves ütemezés	
Oktatási hét (konzultáció)	Témakör
1.	Bevezetés, az adatok jelentősége, az adattudomány gyakorlati alkalmazásai. labor: Az adatok szerepének bemutatása a modern világban, az adattudomány alkalmazási területeinek áttekintése, valamint az alapvető programozási és adatelemzési eszközök megismerése
2.	Az általános elemző adattudományi folyamat, CRISP-DM módszertan. labor: Relációs adatbázisok kezelése és alapvető adatkísérletek.
3.	A különböző adattípusok tulajdonságai, azok gyakorlati szerepe. Alapvető adat előfeldolgozási módszerek. labor: Adatgyűjtés problémái: hiányzó adatok, duplikátumok, hibás formátumok kezelése
4.	Felügyelt gépi tanuló modellek. Regresszió. labor: Statisztikai mutatók számítása és a különböző ML-típusok illeszkedésének vizsgálata példa feladatokon.
5.	Döntési fák. labor: Egyszerű adatelemzési modellek (pl. lineáris regresszió, döntési fák, k-means) megvalósítása és eredményeik értelmezése
6.	Neurális hálók jellemzői. labor: Egyszerű neurális háló implementációja és kiértékelése képfeldolgozási példán
7.	Esettanulmányok labor: 1. zh
8.	Nem felügyelt gépi tanulás. labor: változatos adatkészleteken különböző klaszterező eljárások alkalmazása, klaszterminőség értékelése
9.	Szupport vektor gépek. labor: Szupport vektor gépek
10.	SZÜNET

11.	Modellértékelési szempontok a felügyelt gépi tanulásban, teljesítménymérő metrikák labor: különböző osztályozó és regressziós prediktív modellek teljesítményének kiértékelése valós adatkészleteken, modellek általánosító képességének vizsgálata
12.	Asszociációs szabálybányászat labor: gyakori elemkombinációk és ezek közötti asszociációs szabályok származtatása tranzakciós adatokban (pl. piackosár-elemzés)
13.	zh (labor és elmélet), projektfeladat bemutatás
14.	Pót zh (labor és elmélet)
Félévközi követelmények	
Évközi jegy / aláírás megszerzésének feltételei:	évközi jegy
Zárthelyi dolgozatok	
Oktatási hét	Témakör
7.	labor zh
13.	elméleti zh, labor zh és labor projektfeladat benyújtása
14.	elméleti és labor zárthelyi dolgozatok pótlása
Az évközi jegy kialakításának módszere (csak évközi jegyes tárgyak esetében töltendő ki)	
A félévzáró érdemjegy az elmélet anyagából a 13-ik héten írt zárthelyi dolgozat (50 pont) és a laborokból szerezhető pontok alapján kerül kialakításra. Labor pontok: 1. labor zh (Moodle teszt) a 7-ik héten (10 pont) + labor projekt feladat benyújtás és bemutatás a 12-ik és a 13-ik héten (10 pont) + 2. labor zh Moodle teszt a 13-ik héten (10 pont) + jelenlét (20 pont). Letiltás jár a 30%-on túli igazolatlan hiányzás esetén. A maximálisan elérhető pontszám együttesen 100 pont.	
Pótlás módja	
A ZH / évközi jegy / aláírás pótlásának módja:	Pót zárthelyi írható az elmélet anyagából a 14-ik héten. A labor projektfeladat nem pótolható. A labor zh-k (Moodle tesztek) pótolhatók a 14-ik héten.
Vizsga módja (csak vizsgás tantárgy esetében töltendő ki)	
-	
Vizsgajegy kialakítása (csak vizsgás tantárgy esetében töltendő ki)	
-	
Az egyes érdemjegyek ponthatárai:	
Az érdemjegyek alsó ponthatárai a következők: elégséges – 51 pont, közepes – 63 pont, jó – 74 pont, jeles – 85 pont.	
Irodalom	
Kötelező:	– A Moodle rendszerben közzétett jegyzet.
Ajánlott:	– J. Leskovec, A. Rajaraman, J. D. Ullman: Mining of Massive Datasets – J. Gareth et al.: An Introduction to Statistical Learning
Egyéb segédletek:	