



Tantárgy kód

BMETE959307

Tantárgy azonosító adatok

1.	A tárgy címe	Statisztikai módszerek az adatbányászatban									
2.	A tárgy angol címe	Statistical methods in data mining									
3.	A tárgy rövid címe	Adatstatisztika	Követelmény	2	+	0	+	0	v	Kredit	3
4.	Ajánlott/kötelező tanulmányi rend										
	vagy	Tantárgy kód 1	Rövid cím 1	Tantárgy kód 2	Rövid cím 2	Tantárgy kód 3	Rövid cím 3				
	4.1	BMETE951213	Mat stat ea.								
	4.2	BMETE951233	Mat stat gy.								
	4.3										
5.	Kizáró tantárgyak										
6.	A tantárgy felelős tanszéke	Sztochasztika Tanszék									
7.	A tantárgy felelős oktatója	Dr. Bolla Marianna	beosztása	egyetemi docens							

Akkreditációs adatok

8.	Akkreditációra benyújtás időpontja	2006.11.16.	Akkreditációs bizottsági döntés időpontja	2006.12.19.
----	------------------------------------	--------------------	---	-------------

Megjegyzések

Csak az űrlap fehéren hagyott mezőibe írjunk és a mezők között a **tabulátor** billentyűvel haladjunk! Ha egy kitöltött mezőből tabulátor billentyűvel lépünk ki, több más mező értéke automatikusan megváltozhat. Egy adott mezőre lépve, az állapotsorban megjelenő rövid, ill. az F1 gomb megnyomásakor kapható hosszabb leírás ad segítséget a kitöltéshez. A *tantárgy kódot* a dékáni hivatal adja.

1-2. sorok: A tárgy címének (max. 60 karakter) legalább egy karakterben különböznie kell minden más, Neptunban regisztrált tárgy címétől.

3. sor: A rövid cím jellegzetes, legfeljebb 16 karakter hosszúságú rövidítés. A követelmény eladás+gyakorlat+labor formátumú, az utolsó mező a félév végi számonkérés típusa (v,f,a vagy s, részletes információ az F1 gombra). A kredit megadásánál ügyelni kell arra, hogy az alább részletezett, a tantárgy elvégzéséhez átlagosan szükséges tanulmányi munka mennyiségével összhangban legyen (összes óraszám = kredit*30 óra).

4. sor: Legfeljebb 3, már korábban hallgatott tárgy adható meg a 4.1 sorban. A 4.2 és 4.3 sorok *vagyilag* lehetőségek megadására szolgálnak, például abban az esetben, ha az egyik tárgynak korábban oktatott változatai is megfelelnek. **5. sor:** A *kizáró tantárgyaknál* azokat a tárgyakat kell felsorolni, amelyek tematikái a most akkreditálandó tárggyal 75% vagy annál nagyobb átfedést mutatnak.

6-7. sorok: A felelős tanszék és oktató hatáskörét, ill. kijelölésének feltételeit a *Képzési Kódex 2001* c. dokumentum 9.1 fejezete tartalmazza.

Tematika			
7.	A tantárgy az alábbi témakörök ismeretére épít		
	Paraméteres és nemparaméteres statisztikai módszerek, feltételes várható érték, gráfok, mátrixok.		
8.	A tantárgy célkit zése, feladata a szakképzés céljának megvalósításában		
	Szabadon választható tárgy. A statisztika modern algoritmikus modelljeinek alkalmazása hálózatokra és nagyméret diszkrét adatrendszerre.		
9.	A tantárgy részletes tematikája		
	<p>Napjaink nagyméret , sokdimenziós adatrendszerei a klasszikus statisztikai módszerek keretein túlmutató vizsgálatokat igényelnek. Bemutatjuk az erre a célra az utóbbi ötven évben kifejlesztett ún. algoritmikus modelleket: EM-algoritmus hiányos adatokra, ACE-algoritmus általánosított regresszióra, Kaplan-Meier becslések cenzorált adatokra, jackknife és bootstrap algoritmusok újramintavételezésre. Regresszió és diszkriminanciaanalízis összekapcsolása a Gauss-Markov elmélet és Bayes-döntések alapján. Autokorrelációk vizsgálata id sorokban, ARMA-folyamatok. Paraméterbecslés, identifikáció.</p> <p>Hálózatok struktúrájának feltárása nemparaméteres statisztikai módszerek és gráfok spektrális tulajdonságai segítségével. Többváltozós korrespondanciaanalízis és homogenitásvizsgálat többdimenziós kontingenciatáblákra, diszkrét változók varianciaanalízise, rangstatisztikák. Gráfok és hipergráfok Laplace-mátrixa, k-vágások mér számainak becslése a sajátértékekkel. Spektrális gráfklaszterezés a Laplace-mátrix sajátvektoraival. Nagyméret véletlen mátrixok kiugró sajátértékeinek ill. szinguláris értékeinek aszimptotikus viselkedése és eloszlásuk speciális esetekben. Véletlenített módszerek nagyméret téglalpmátrixok szinguláris felbontására. Hatvány fokszámeloszlású véletlen gráfok statisztikus jellemz i.</p> <p>A matematikai egzaktsággal tárgyalt elméleten túl szó lesz konkrét alkalmazási lehet ségekr l az adatbányászatban (internetes, pénzügyi adatrendszerek, kommunális hálózatok, microarray, sejtek metabolikus rendszere).</p>		
10.	Követelmények, az osztályzat (aláírás) kialakításának módja		
	szorgalmi id szakban	szabadon feldolgozott cikk, órák látogatása	vizsgaid szakban vizsga
11.	Pótlási lehet ségek		
	vizsgaid szak els hetében pótbeszámoló a feldolgozott cikkr l		
12.	Konzultációs lehet ségek		
	hetente fogadóóra, vizsgák el tt külön konzultáció		
13.	Jegyzet, tankönyv, felhasználható irodalom		
	T. Hastie, R. Tibshirani, J. Friedman: The Elements of Statistical Learning. Data Mining... Springer, New York, 2001.		
	B. Bollobás: Random Graphs, 2 nd ed., Cambridge University Press, 2001.		
	F. R. K. Chung: Spectral Graph theory, CBMS Series, No. 92, AMS Publ., 1997.		

14.	A tantárgy elvégzéséhez átlagosan szükséges tanulmányi munka mennyisége órákban (a teljes szemeszterre számítva)		
	14.1	Kontakt óra	28
	14.2	Félévközi felkészülés órákra	28
	14.3	Felkészülés zárthelyire	0
	14.4	Zárthelyik megírása	0
	14.5	Házi feladat elkészítése	12
	14.6	Kijelölt írásos tananyag elsajátítása (beszámoló)	5
	14.7	Egyéb elfoglaltság	0
	14.8	Vizsgafelkészülés	17
	14.9	Összesen	90
15.	Ellenrz adat		Kredit * 30

A tantárgy tematikáját kidolgozta			
16.	Név	beosztás	Munkahely (tanszék, kutatóintézet stb.)
	Dr. Bolla Marianna	egyetemi docens	Sztochasztika Tanszék

A tanszékvezet		
17.	Neve	aláírása
	Dr. Tóth Bálint	

Megjegyzések

14.1 sor: Értéke automatikusan kitölt dik az rlap elektronikus változatában, a „Követelmény” címszónál megadott óraszám értékek alapján, az (el adás+gyakorlat+labor) * (14 oktatási hét) formula szerint. **14.4 sor:** Értéke 0, ha a zárthelyik íratása kontakt órákon történik, egyébként pedig a minimálisan szükséges számú zárthelyi megírásához felhasználandó id (a pót zárthelyik nélkül). **14.7 sor:** Az „Egyéb elfoglaltság” szöveg helyére a tevékenység konkrét megnevezését kell írni.

15. sor: Az itt szerepl értéknek és a **14.9 sorban** automatikusan megjelen tanulmányi óraszám összegnek hozzávet legesen meg kell egyeznie! Tájékoztatásul azt vegyük figyelembe, hogy a hallgatók által egy szemeszterben átlagosan 30 kreditnyi munkamennyiséget kell teljesíteni, azaz a szorgalmi és vizsgaid szak során elvárt terhelés összesen kb. 900 munkaóra.