

INFORMAČNÝ LIST PREDMETU

Vysoká škola: Univerzita Konštantína Filozofa v Nitre	
Fakulta: Fakulta prírodných vied	
Kód predmetu: KI/ DM/15	Názov predmetu: Data mining
Druh, rozsah a metóda vzdelávacích činností: Forma výučby: Prednáška / Cvičenie Odporúčaný rozsah výučby (v hodinách): Týždenný: 2 / 2 Za obdobie štúdia: 26 / 26 Metóda štúdia: prezenčná	
Počet kreditov: 6	
Odporúčaný semester/trimester štúdia: 3.	
Stupeň štúdia: II.	
Podmieňujúce predmety:	
Podmienky na absolvovanie predmetu: Úspešné absolvovanie predmetu je podmienené skúškou, samostatným riešením projektu a úloh zadaných v kurze. Ústna skúška pozostáva z teoretických otázok (30%), z otázok týkajúcich sa riešených úloh v kurze (30%) a z obhajoby projektu (40%). Hodnotenie predmetu je dané výsledkom skúšky. Hodnotenie: A - výborne: 100% - 93% B - veľmi dobre: 92,99% - 85% C - dobre: 84,44% - 77% D - uspokojivo: 76,99% - 69% E - dostatočne: 68,99% - 60% FX - nedostatočne: 59,99% - 0%	
Výsledky vzdelávania: Študent rozumie procesu objavovania znalostí- je oboznámený s procesom transformácie údajov na informácie a znalosti. Dôraz je kladený na modelovanie (Data Mining, DM), ktoré je jadrom celého procesu objavovania znalostí (Knowledge Discovery, KD). Študent diskutuje o možnostiach analýzy dát- aplikácii analytických metód na riešenie konkrétnych data mining-ových problémov/ úloh objavovania znalostí. Pri riešení problémov sa kladie dôraz nie na algoritmy výpočtov, ale na podstatu metód, správny výber metódy a vyhodnotenie výsledkov (overenie predpokladov použitia, porovnanie výsledkov, vizualizácia, krížová validácia). Študent po absolvovaní predmetu je schopný analyzovať predspracované dáta, vybrať a aplikovať správnu metódu na riešenie data mining-ového problému/úlohy objavovania znalostí a vyhodnotiť získané výsledky.	
Stručná osnova predmetu: 1. Oblasti objavovania znalostí (KDD, text mining, web mining); zdroje dát (databázy/dátové sklady, kolekcie dokumentov, logovacie súbory). 2. Úlohy objavovania znalostí/typy DM problémov (deskripcia dát a sumarizácia, segmentácia, deskripcia konceptov, klasifikácia, predikcia, analýza závislostí). 3. Dolovanie z dát; výber analytickej metódy; štatistické metódy (základné štatistické metódy, viacrozmerne prieskumné techniky, lineárne/nelineárne modely); metódy strojového učenia (symbolické, subsymbolické metódy); aplikácia vybraných analytických metód; vyhodnotenie	

výsledkov (overenie predpokladov použitia, porovnanie výsledkov, vizualizácia, krížová validácia, m-násobná krížová validácia);

4. Projekt: metodika modelovania správania sa používateľov webu v závislosti od času (určenie modelu, odhad parametrov modelu maximalizáciou logaritmu multinominálnej funkcie vierohodnosti, odhad logitov pre všetky hodnoty nezávislých premenných, odhad pravdepodobností prístupov, vizualizácia pravdepodobností výberu webovej časti j v čase i, evalvácia modelu, porovnanie empirických a teoretických početností/pravdepodobností/logitov, analýza rezíduí, krížová validácia).

Odporúčaná literatúra:

- Berka, P. 2003. Dobývání znalostí z databází. Praha : Academia, 2003. ISBN 80-200-1062-9.
- Hill, T. - Lewicki, P. 2006. Statistics: methods and applications : a comprehensive reference for science, industry and data mining. Tulsa : StatSoft, 2006. 832 p. ISBN 1-884233-59-7.
- Chapman, P. et al. 2000. CRISP-DM 1.0 Step-by-step data mining guide. SPSS, 2000.
- Liu, B. 2007. Web data mining: Exploring hyperlinks, contents and usage data. Springer, 2007. ISBN 978-3-540-37881-5.
- Munk, M. - Drlík, M. - Vrábelová, M. 2011. Probability Modelling of Accesses to the Course Activities in the Web-based Educational System. In Lecture Notes in Computer Science. Springer, ISSN 0302-9743, 2011, vol. 6786, p. 485-499.
- Munk, M. - Drlík, M. 2014. Analysis of Stakeholders' Behaviour Depending on Time in Virtual Learning Environment. In Applied Mathematics & Information Sciences. ISSN 1935-0090, 2014, vol. 8, no. 2, p. 773-785.
- Munk, M. - Pilková, A. - Kapusta, J. - Švec, P. - Drlík, M. 2013. Pillar 3 and Modelling of Stakeholders Behaviour at the Commercial Bank Website during the Recent Financial Crisis. In Procedia Computer Science. Elsevier, ISSN 1877-0509, 2013, vol. 18, p. 1747-1756.
- Munk, M. - Vrábelová, M. - Kapusta, J. 2011. Probability Modeling of Accesses to the Web Parts of Portal. In Procedia Computer Science. Elsevier, ISSN 1877-0509, 2011, vol. 3, p. 677-683.
- Munk, M. 2011. Počítačová analýza dát. Nitra : UKF, 2011. 361 s. ISBN 978-80-8094-895-5.
- Paralič, J. 2003. Objavovanie znalostí v databázach. Košice : Elfa, 2003. 80 s. ISBN 80-89066-60-7.
- Paralič, J. a kol. 2010. Dolovanie znalostí z textov. Košice : Equilibria, 2010. 184 s.

Jazyk, ktorého znalosť je potrebná na absolvovanie predmetu:

slovenský

Poznámky:

Hodnotenie predmetov

Celkový počet hodnotených študentov: 298

A	B	C	D	E	FX
59.4	10.74	11.74	8.72	7.72	1.68

Vyučujúci: prof. RNDr. Michal Munk, PhD.

Dátum poslednej zmeny: 17.02.2020

Schválil: prof. RNDr. Michal Munk, PhD.