

UČNI NAČRT PREDMETA / COURSE SYLLABUS						
Predmet:		Odkrivanje znanj iz podatkov				
Course title:		Data mining				
Študijski program in stopnja Study programme and level		Študijska smer Study field		Letnik Academic year	Semester Semester	
Interdisciplinarni magistrski študijski program Računalništvo in matematika		ni smeri		1 in 2	drugi	
Interdisciplinary Masters study programme Computer Science and Mathematics		none		1 in 2	second	
Vrsta predmeta / Course type				izbirni		
Univerzitetna koda predmeta / University course code:				63525		
Predavanja Lectures	Seminar Seminar	Vaje Tutorial	Klinične vaje work	Druge oblike študija	Samost. delo Individ. work	ECTS
45	20	10			105	6
Nosilec predmeta / Lecturer:		Blaž Zupan				
Jeziki / Languages:	Predavanja / Lectures:	slovenski/Slovene, angleški/English				
	Vaje / Tutorial:	slovenski/Slovene, angleški/English				
Pogoji za vključitev v delo oz. za opravljanje študijskih obveznosti:				Prerequisites:		
Vpis predmeta Uvod v odkrivanje znanj iz podatkov iz predhodnih semestrov.						
Vsebina:				Content (Syllabus outline):		

<p>Predmet bo v teoriji in na praktičnih primerih predstavil sledeče vsebine:</p> <ol style="list-style-type: none"> 1. Predstavitev področja in klasifikacija tehnik za odkrivanje znanj iz podatkov, pregled značilnih aplikacij 2. Tehnološke platforme in razvojne metodologije (skriptna okolja, okolja za analizo podatkov z vizualnim programiranjem) 3. Predobdelava podatkov: iskanje osamelcev, zmanjševanje dimenzij (metoda glavnih komponent), izbor in konstrukcija značilnik, permutacijski pristopi, diskretizacija 4. Uvrščanje v skupine, s poudarkom na tehnikah, ki lahko obravnavajo velike množice podatkov in podatkov z velikim naborom značilnik, metode podpornih vektorjev, iskanje in vizualizacija interakcij 5. Tehnike razvrščanja v skupine (metode hierarhičnega združevanja, metode voditeljev), s poudarkom na tehnikah, ki lahko obravnavajo velike množice podatkov, določanje števila skupin (metoda silhuete) 6. Ocenjevanje uspešnosti napovednih modelov, kalibracijske in diskriminantne metode, ROC analiza, permutacijski pristopi 7. Vizualizacija podatkov in modelov, tehnike gradnje, analize in vizualizacije mrež 8. Tehnike odkrivanja znanj iz zbirk besedil in spletnih strani 9. Integrativni pristopi (uporaba predznanja, integracija povezav, pridobljenih iz različnih naborov podatkov) 10. Tipične napake pri snovanju pristopov ali uporabi tehnik odkrivanja znanj iz podatkov in 	<p>The course will cover theoretical and practical aspects of the following data mining approaches:</p> <p>Introduction to data mining, taxonomy of data mining approaches and tasks</p> <p>Data mining programming environments (scripting, visual programming)</p> <p>Data preprocessing (dimensionality reduction, feature construction, identification of outliers)</p> <p>Classification, including support vector machines and feature interaction discovery</p> <p>Clustering, with emphasis on techniques that can consider very large data sets, and techniques for to determine an appropriate number of clusters</p> <p>Evaluation, including permutation-based and cross-validation approaches, statistical scoring of models</p> <p>Data and model visualization techniques, visualization of networks</p> <p>Text mining, text-based kernels for support vector machines</p> <p>Integrative aspects, including ensemble methods and mining with inclusion of prior knowledge</p> <p>Typical mistakes in data mining and how to avoid them</p> <p>The course will be composed of lectures in core data mining techniques and tools, which will then be employed on practical problems during lab work. We will focus on open source solutions and modern scripting languages (e.g., Python). Students will use scripting to access various data mining techniques which they, in a programming framework, will combine into</p>
---	--

<p>kako se jim izognemo</p> <p>Na predavanjih bodo študenti spoznavali ključne tehnologije in orodja, s katerimi bodo tekom semestra na vajah in v okviru projektov oz. seminarskih nalog reševali praktične probleme. Poudarek bo na uporabi odprtokodnih, prosto dostopnih orodij, ki za analizo podatkov uporabljajo moderne skriptne jezike (npr. Python). V skriptnih okoljih bodo študenti z uporabo že obstoječih komponent razvijali lastne metode, uporabo teh preverjali na različnih podatkih, ter poročali o ocenah njihove uporabnosti in napovedne točnosti. Vaje se bodo izvajale v računalniški učilnici opremljeni z ustrežno strojno in programsko opremo.</p>	<p>their own data mining procedures.</p>
--	--

Temeljni literatura in viri / Readings:

1. Tan P-N, Steinbach M, Kumar V (2006) Introduction to data mining. Pearson Education, Boston.
2. Leskovec J, Rajaraman A, Ullman J (2014) Mining of Massive Datasets, Cambridge University Press, 2 edition.
3. Chollet F (2018) Deep learning with Python, Manning Publications.

Cilji in kompetence:

Cilj predmeta je študente seznaniti z osnovnimi in naprednimi metodami odkrivanja znanj iz podatkov, s poudarkom na njihovi praktični uporabi. Pri predmetu se bodo naučili uporabljati moderna skriptna orodja za analizo podatkov. Spoznali bodo, kako je z njimi moč implementirati nove metode za odkrivanje znanj, oziroma kako je moč obstoječe tehnike prilagoditi za obravnavo konkretnih podatkov.

Objectives and competences:

Students will learn a number of core techniques for data mining. The course will include an introduction to data mining as well as a detailed study of several selected methods. It will also focus on practical use of these methods on real-life problems. The course will use a scripting data mining environment, where students will learn how to use the existing data mining libraries and design and implement in code their own data mining solutions.

Predvideni študijski rezultati:

Intended learning outcomes:

<p>Po uspešnem zaključku predmeta bo študent: sposoben prepoznati probleme, kjer bi si pri rešitvi pomagal s tehnikami strojnega učenja,</p> <p>sposoben predstaviti problemsko znanje oziroma podatke v obliki primerne za strojno učenje,</p> <p>razumel razliko med različnimi tehnikami odkrivanja znanj iz podatkov,</p> <p>iz problema in podatkov prepoznal priložnosti in koristi uporabe posameznih tehnik podatkovne analitike,</p> <p>sposoben izdelati program za analizo podatkov in uporabo knjižnic za podatkovno analitiko v jeziku Python,</p> <p>razumel uporabo knjižnic za globoko učenje,</p> <p>sposoben samostojno uporabiti knjižnice za globoko učenje,</p> <p>razumel matematične osnove glavnih metod podatkovne analitike.</p>	<p>After the completion of the course the student will be able to:</p> <p>recognize problems where one can apply machine learning,</p> <p>understand the process of transformation of the problem-specific data to the form suitable for data mining,</p> <p>understand the difference of various techniques of data mining in application to the real-world data,</p> <p>identify what kind of advantages of different machine learning techniques provide for specific data sets,</p> <p>be able to write Python scripts for data analytics and within them integrate various data mining libraries,</p> <p>use libraries for deep learning,</p> <p>understand the mathematics behind most of the data mining approaches.</p>

Metode poučevanja in učenja:

Predavanja s podporo avdio-vizualne opreme, sprotni razvoj programskih rešitev, laboratorijske vaje v računalniški učilnici z ustrezno programsko opremo. Delo posamezno in v skupinah. Velik poudarek na praktičnem delu (npr. razvoj skript za pregledovanje in analizo podatkov) in reševanju praktičnih problemov.

Learning and teaching methods:

Combined lecturing with simultaneous use of the blackboard and computer projection (coding, visualization of models, results). Lab work in computer-equipped lecture rooms. Individual and work in team. Emphasis on practical problem solving.

Delež (v %) /

Načini ocenjevanja:

Weight (in %) **Assessment:**

<p>Način (pisni izpit, ustno izpraševanje, naloge, projekt): Sprotno preverjanje (domače naloge, kolokviji in projektno delo)</p> <p>Končno preverjanje (pisni in ustni izpit)</p> <p>Ocene: 6-10 pozitivno, 5 negativno (v skladu s Statutom UL).</p>	<p>50%</p> <p>50%</p>	<p>Type (examination, oral, coursework, project): Continuing (homework, midterm exams, project work)</p> <p>Final (written and oral exam)</p> <p>Grading: 6-10 pass, 5 fail (according to the rules of University of Ljubljana).</p>
--	-----------------------	--

Reference nosilca / Lecturer's references:

Blaž Župan:

- ZUPAN, Blaž, DEMŠAR, Janez, BRATKO, Ivan, JUVAN, Peter, HALTER, John A., KUSPA, Adam, SHAULSKY, Gad. GenePath : a system for automated construction of genetic networks from mutant data. *Bioinformatics*, ISSN 1367-4803. [Print ed.], 2003, vol. 19, no. 3, str. 383-389 [COBISS.SI-ID 3415124]
- VAN DRIESSCHE, Nancy, DEMŠAR, Janez, BOOTH, Egzi O., HILL, Paul, JUVAN, Peter, ZUPAN, Blaž, KUSPA, Adam, SHAULSKY, Gad. Epistasis analysis with global transcriptional phenotypes. *Nature genetics*, ISSN 1061-4036, May 2005, vol. 37, no. 5, str. 471-477, ilustr [COBISS.SI-ID 4712532]
- MRAMOR, Minca, LEBAN, Gregor, DEMŠAR, Janez, ZUPAN, Blaž. Visualization-based cancer microarray data classification analysis. *Bioinformatics*, ISSN 1367-4803. [Print ed.], 2007, vol. 23, no. 16, str. 2147-2154, ilustr [COBISS.SI-ID 6087252]
- BELLAZZI, Riccardo, ZUPAN, Blaž. Predictive data mining in clinical medicine : current issues and guidelines. *International journal of medical informatics*, ISSN 1386-5056. [Print ed.], 2008, vol. 77, no. 2, str. 81-97, ilustr [COBISS.SI-ID 6280788]
- ZUPAN, Blaž, BOHANEK, Marko, DEMŠAR, Janez, BRATKO, Ivan. Learning by discovery concept hierarchies. *Artificial intelligence*, ISSN 0004-3702. [Print ed.], 1999, vol. 109, str. 211-242 [COBISS.SI-ID 14228007]