UČNI NAČRT PREDMETA / COURSE SYLLABUS	
Predmet:	Računalniško podprto odkrivanje znanstvenih zakonitosti iz strukturiranih, prostorskih in časovnih podatkov
Course title:	Computational Scientific Discovery from Structured, Spatial and Temporal Data

Študijski program in stopnja **Study programme and level**	**Modul** **Module**	**Letnik** **Academic year**	**Semester** **Semester**
Informacijske in komunikacijske tehnologije, 3. stopnja	Tehnologije znanja	1	1
Information and Communication Technologies, 3rd cycle	Knowledge Technologies	1	1

Vrsta predmeta / Course type Izbirni / Elective

Univerzitetna koda predmeta / University course code: IKT3-723

Predavanja **Lectures**	**Seminar** **Seminar**	**Sem. vaje** **Tutorial**	**Lab. vaje** **Laboratory work**	**Druge oblike**	**Samost. delo** **Individ. work**	**ECTS**
15	15			15	105	5

**Navedena porazdelitev ur velja, če je vpisanih vsaj 15 študentov. Drugače se obseg izvedbe kontaktnih ur sorazmerno zmanjša in prenese v samostojno delo. / This distribution of hours is valid if at least 15 students are enrolled. Otherwise the contact hours are linearly reduced and transfered to individual work.*

Nosilec predmeta / Lecturer: Prof. dr. Sašo Džeroski

Jeziki / Languages:		
	Predavanja / Lectures:	Slovenščina, angleščina / Slovenian, English
	Vaje / Tutorial:	

Pogoji za vključitev v delo oz. za opravljanje študijskih obveznosti:	**Prerequisites:**
Zaključen študij druge stopnje s področja informacijskih ali komunikacijskih tehnologij ali zaključen študij druge stopnje na drugih področjih z znanjem osnov s področja predmeta. Potrebna so tudi osnovna znanja matematike, računalništva in informatike.	Completed second cycle studies in information or communication technologies or completed second cycle studies in other fields with knowledge of fundamentals in the field of this course. Basic knowledge of mathematics, computer science and informatics is also requested.

Vsebina:	**Content (Syllabus outline):**
Različne naloge napovedovanja strukturiranih vrednosti: večciljna klasifikacija in regresija, (hierarhična) večznačna klasifikacija, napovedovanje kratkih časovnih vrst. Dodatne dimenzije kompleksnosti: nepopolne označbe, podatkovni tokovi in omrežni podatki. Napovedno razvrščanje za napovedovanje strukturiranih vrednosti: Uvod v napovedno razvrščanje, drevesa za napovedno razvrščanje za različne tipe ciljnih vrednosti, učenje	The different tasks of predicting structured outputs: multi-target classification and regression; (hierarchical) multi-label classification; time-series as targets. Additional dimensions of complexity: incomplete annotations, streaming and network data. Predictive clustering for structured output prediction: Introduction to predictive clustering, predictive clustering trees for different targets, constraint-based learning

tovrstnih dreves z omejitvami.
Ontologije za podatkovno rudarjenje: Ontologija podatkovnih tipov, ontologija ključnih pojmov podatkovnega rudarjenja, opis napovedovanja strukturiranih vrednosti.
Ansambelske metode za napovedovanje strukturiranih vrednosti: Ansambli dreves, ansambli pravil, rangiranje značilk.
Napredne teme: Pol-nadzorovano učenje za napovedovanje strukturiranih vrednosti, učenje iz podatkovnih tokov.
Primeri uporabe napovedovanja strukturiranih vrednosti: Znanosti o okolju, napovedovanje zgradbe združb, znanosti o življenju, npr. napovedovanje funkcij genov.

thereof.
Ontologies for data mining: Ontology of data types, ontology of core data mining entities, describing structured output prediction.
Ensemble methods for structured output prediction: Tree ensembles, rule ensembles, feature ranking.
Advanced topics: Semi-supervised learning for structured-output prediction, structured output prediction on data streams
Applications of structured output prediction: Environmental sciences (ecology, e.g., predicting community structure), life sciences (systems biology, e.g., predicting gene function), image annotation and retrieval.

Temeljna literatura in viri / Readings:

Izbrana poglavja iz naslednjih knjig: / Selected chapters from the following books:

- S. Džeroski, B. Goethals, and P. Panov, Eds. *Inductive Databases and Constraint-Based Data Mining*. Springer, 2010. ISBN 978-1-4419-7737-3
- X. Zhu, and A. Goldberg. Introduction to Semi-Supervised Learning. Morgan and Claypool, 2009. ISBN 978-1-5982-9547-4
- S. Džeroski, B. Goethals, and P. Panov, Eds. Inductive Databases and Constraint-Based Data Mining. Springer, 2010. ISBN 978-1-4419-7737-3.
- V. Bolon-Canedo, N. Sanchez-Marono, and A. Alonso-Betanzos. Feature Selection for High-Dimensional Data, Springer, 2016. ISBN 978-3-3192-1857-1.
- F. Herera, F. Charte, A. Rivera, and M. del Jesus. Multilabel Classification: Problem Analysis, Metrics and Techniques. Springer, 2016. ISBN 978-3-3194-1110-1.
- A. Bifet, R. Gavalda, B. Pfahringer, and G. Holmes. Machine Learning for Data Streams: with Practical Examples in MOA. MIT Press, 2018. ISBN 978-0-2620-3779-2

Cilji in kompetence:

Cilj predmeta je seznaniti študenta s področjem računalniškega odkrivanja znanstvenih zakonitosti iz kompleksnih podatkov, vključno s strukturiranimi, prostorskimi in časovnimi podatki, s poudarkom na napovedovanju strukturiranih vrednosti.

Kompetence študenta z uspešno zaključenim predmetom bodo vključevale razumevanje osnovnih nalog odkrivanja znanja iz tega področja, poznavanje sodobnih metod za reševanje takih nalog ter znanje o primerih uporabe le-teh na dveh pomembnih znanstvenih področjih (znanosti o okolju in znanosti o življenju).

Objectives and competences:

The goal of the course is to familiarize the student with the field of computational scientific discovery from complex data, including structured, spatial and temporal data and in particular predicting structured outputs.

The competencies of the students completing this course successfully would include understanding of basic tasks from the area, familiarity with state-of-the art methods for solving them, and knowledge of example applications of these methods in two major scientific fields (environmental and life sciences).

Predvideni študijski rezultati:

Študenti bodo z uspešno opravljenimi obveznostmi tega predmeta pridobili veščine in sposobnosti uporabe metod strojnega učenja za:

- večciljno klasifikacijo in regresijo
- (hierarhično) večznačno klasifikacijo
- napovedovanje kratkih časovnih vrst
- pol-nadzorovane različice zgornjih nalog
- različice zgornjih nalog, kjer se je potrebno učiti iz podatkovnih tokov

Spoznali bodo in se naučili uporabljati ontologije podatkovnega rudarjenja, in sicer:

- ontologijo podatkovnih tipov in
- ontologijo ključnih pojmov podatkovnega rudarjenja

tako za iskanje kot za označevanje algoritmov in podatkov.

Pridobili bodo tudi sposobnosti ugotoviti, če in katere metode računalniškega odkrivanja znanstvenih zakonitosti iz kompleksnih podatkov je potrebno uporabiti za analizo dane množice znanstvenih podatkov.

Intended learning outcomes:

Students successfully completing this course will acquire skills and capabilities of using machine learning methods for:

- multi-target classification and regression
- (hierarchical) multi-label classification
- predicting short time series
- semi-supervised learning variants of the above tasks
- data stream learning variants of the above tasks

They will also get familiar with and be able to use ontologies for data mining

- ontology of data types
- ontology of data mining entities

for finding and annotating algorithms and data

They will acquire the ability to identify whether and which methods for computational scientific discovery from complex data are needed to analyse a given set of scientific data.

Metode poučevanja in učenja:

Predavanja, konzultacije, individualno delo

Learning and teaching methods:

Lectures, consultancy, individual work

Načini ocenjevanja:	Delež (v %) / Weight (in %)	**Assessment:**
Seminarska naloga	50 %	Seminar work
Ustni zagovor seminarske naloge	50 %	Oral defense of seminar work

Reference nosilca / Lecturer's references:

- J. Levatić, D. Kocev, M. Ceci, **S. Džeroski**. Semi-supervised trees for multi-target regression. Information Sciences 450: 109-127, 2018.
- A. Osojnik, P. Panov, **S. Džeroski**. Multi-label classification via multi-target regression on data streams. Machine Learning 106 (6), 745-770, 2017.
- J. Levatić, D. Kocev, **S. Džeroski**. The importance of the label hierarchy in hierarchical multi-label classification. Journal of Intelligent Information Systems. 45 (2), 247-271, 2015.
- P. Panov, L. Soldatova, and **S. Džeroski.** Ontology of core data mining entities. Data Mining and Knowledge Discovery 28 (5-6), 1222-1265, 2014.
- D. Kocev, C. Vens, J. Struyf, and **S. Džeroski.** Tree ensembles for predicting structured outputs. Pattern Recognition 46 (3), 817-833, 2013.