

UČNI NAČRT PREDMETA / COURSE SYLLABUS

Predmet:	Podatkovno rudarjenje in odkrivanje zakonitosti
Course title:	Data Mining and Knowledge Discovery

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, 3 rd cycle	Knowledge Technologies	1	1

Vrsta predmeta / Course type

Izbirni / Elective

Univerzitetna koda predmeta / University course code:

IKT3-722

Predavanja Lectures	Seminar Seminar	Sem. vaje Tutorial	Lab. vaje Laboratory work	Druge oblike	Samost. delo Individ. work	ECTS
30	30			30	210	10

**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 transferred to individual work.*

Nosilec predmeta / Lecturer:

Prof. dr. Nada Lavrač
Prof. dr. Bojan Cestnik
Doc. dr. Martin Žnidaršič
Doc. dr. Petra Kralj Novak

**Jeziki /
Languages:**

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

Pogoji za vključitev v delo oz. za opravljanje študijskih obveznosti:

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.

Prerequisites:

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:

Uvod:
uvod v podatkovno rudarjenje in odkrivanje zakonitosti v podatkih, povezava s strojnim učenjem, vizualizacija podatkov in modelov, predstavitev CRISP-DM metodologije odkrivanja zakonitosti
Priprava podatkov:
tabelarni podatki in relacijske podatkovne baze, obravnava manjkajočih in šumnih

Content (Syllabus outline):

Introduction:
introduction to data mining and knowledge discovery in databases, relation with machine learning, visualization of data and models, presentation of the CRISP-DM knowledge discovery methodology
Data preparation and preprocessing:
tabular data and relational databases, handling of missing and noisy values,

vrednosti, izbor atributov in značilk

Tehnike rudarjenja podatkov:
 predstavitev posameznih tehnik rudarjenja podatkov: učenje odločitvenih, regresijskih in modelnih dreves, učenje klasifikacijskih in povezovalnih pravil, razvrščanje v skupine, metoda najbližjih sosedov, Naivni Bayesov klasifikator, metoda podpornih vektorjev, umetne nevrnske mreže, odkrivanje podskupin, ansambli klasifikatorjev

Hevristike in ocenjevanje rezultatov:
 predstavitev preiskovalnih hevristik, hevristik za ocenjevanje kvalitete naučenih vzorcev in modelov, metodologija evalvacije rezultatov

Napredne metode rudarjenja podatkov:
 Učenje z delno-označenimi podatki, aktivno učenje, rudarjenje relacijskih podatkov, propozicionalizacija, semantično rudarjenje podatkov

Praktično usposabljanje:
 praktična uporaba izbranih tehnik in orodij rudarjenja podatkov

attribute/feature subset selection

Data mining techniques:
 presentation of specific data mining techniques: decision, regression and model tree learning, learning classification and association rules, clustering, nearest neighbors approach, Naive Bayesian classifier, support vector machines, artificial neural networks, subgroup discovery, ensemble classifiers

Heuristics and results evaluation:
 presentation of search heuristics, heuristics for estimating the quality of induced patterns and models, methodology for results evaluation

Advanced data mining methods:
 Semi-supervised learning, active learning, relational data mining, propositionalization, semantic data mining

Practical training:
 practical use of selected data mining techniques and tools

Temeljna literatura in viri / Readings:

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

- J.H. Witten, E. Frank, and M.A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques (Third Edition)*. Morgan Kaufmann, 2011. ISBN 978-0-12-374856-0
- T. Mitchell, *Machine Learning*. McGraw Hill, 1997. ISBN 0070428077
- M. Berthold, and D.J. Hand, Eds. *Intelligent Data Analysis: An Introduction*. Springer, 2003. ISBN 978-3-540-43060-5
- S. Džeroski, and N. Lavrač, Eds. *Relational Data Mining*. Springer, 2001. ISBN 3-540-42289-7
- J. Fürnkranz, D. Gamberger, and N. Lavrač, *Foundations of Rule Learning*. Springer, 2012. ISBN 978-3-540-75196-0
- M. Bramer, *Principles of Data Mining*. Springer, 2007. ISBN 978-1-84628-765-7

Cilji in kompetence:

Odkrivanje zakonitosti v podatkih je proces odkrivanja vzorcev in modelov, opisanih s pravili ali drugimi človeku razumljivimi formalizmi za predstavitev znanja. Najpomembnejši del tega procesa predstavlja podatkovno rudarjenje, ki vključuje uporabo metod, tehnik in orodij za avtomatsko konstrukcijo vzorcev in modelov iz podatkov.

Cilji predmeta so (a) predstaviti osnove podatkovnega rudarjenja, postopke odkrivanja zakonitosti v podatkih ter metodologijo CRISP-DM, (b) predstaviti izbrane metode in tehnike podatkovnega rudarjenja, (c) predstaviti metodologijo ocenjevanja rezultatov.

Objectives and competences:

Knowledge discovery in databases is a process of discovering patterns and models, described by rules or other human understandable representation formalisms. The most important step in this process is data mining, performed by using methods, techniques and tools for automated constructions of patterns and models from data.

The course objectives are to (a) introduce the basics of data mining, the process of knowledge discovery in databases and the CRISP-DM methodology, (b) present selected data mining methods and techniques, (c) present the methodology for result evaluation.

Študenti bodo obvladali osnove predprocesiranja podatkov, rudarjenja podatkov in odkrivanja zakonitosti v podatkih ter bodo usposobljeni za praktično uporabo izbranih orodij podatkovnega rudarjenja in metod za evalvacijo rezultatov.

The students will master the basics of data preprocessing, data mining and knowledge discovery and will be capable of using selected data mining tools and results evaluation methods in practice.

Predvideni študijski rezultati:

Obvladana uporaba izbranih metod in tehnik podatkovnega rudarjenja, usposobljenost za predprocesiranje podatkov, praktično uporabo izbranih orodij podatkovnega rudarjenja, usposobljenost za uporabo in interpretacijo metod za evalvacijo rezultatov.

Intended learning outcomes:

Mastering of selected data mining methods and techniques, capability of data preprocessing, practical use of selected data mining techniques, and capability of using and interpreting the methods for result evaluation.

Metode poučevanja in učenja:

Predavanja, seminar, konzultacije, individualno delo

Learning and teaching methods:

Lectures, seminar, consultancy, individual work

Načini ocenjevanja:	Delež (v %) / Weight (in %)	Assessment:
Pisni ali ustni izpit	40 %	Written or oral exam
Seminarska naloga	30 %	Seminar work
Ustni zagovor seminarske naloge	30 %	Oral defense of the seminar work

Reference nosilca / Lecturer's references:

- J. Fürnkranz, D. Gamberger, and **N. Lavrač**, *Foundations of Rule Learning*. Springer, 2012.
- A. Vavpetič, V. Podpečan, and **N. Lavrač**, Semantic subgroup explanations. *J. Intell. Inf. Syst.* 42(2): 233-254, 2014.
- Petrič, B. Cestnik, **N. Lavrač**, and T. Urbančič, Outlier detection in cross-context link discovery for creative literature mining. *The Computer Journal* 55(1): 47-61, 2012.
- B. Sluban, D. Gamberger, and **N. Lavrač**, Ensemble-based noise detection: noise ranking and visual performance evaluation. *Data Min. Knowl. Discov.* 28(2): 265-303, 2014.
- M. Grčar, N. Trdin, and **N. Lavrač**. A methodology for mining document-enriched heterogeneous information networks. *The Computer Journal*, 56(3): 321-335, 2013.