

UČNI NAČRT PREDMETA / COURSE SYLLABUS**Predmet:** Uvod v podatkovne znanosti**Course title:** Introduction to Data Science

Študijski program in stopnja	Študijska smer	Letnik	Semester
Study programme and level	Study field	Academic year	Semester

Podatkovne znanosti, magistrski študijski program druge stopnje	-	Prvi	Prvi
The second cycle masters study programme Data Sciences	-	First	First

Vrsta predmeta / Course type

Obvezni / Compulsory

Univerzitetna koda predmeta / University course code:

2-PZ-MAG-UPZ-2020-06-30

Predavanja	Seminar	Vaje	Klinične vaje	Druge oblike študija	Samost. delo	ECTS
Lectures	Seminar	Tutorial	work	work	Individ. work	
30	-	40	-	-	110	6

Nosilec predmeta / Lecturer: Izr. prof. dr. Zoran Levnajić**Jeziki / Predavanja / Languages: Lectures:** Slovenščina/English**Vaje / Tutorial:** Slovenščina/English**Pogoji za vključitev v delo oz. za opravljanje študijskih obveznosti:****Prerequisites:**

Za vključitev v delo mora študent poznati osnovne principe programiranja (v poljubnem programskem jeziku). Zahteva se tudi poznavanje osnov matematike in statistike.

Students need basic familiarity with computer programming (in any programming language). Also, they need solid background in undergraduate mathematics and statistics.

Vsebina:

Uvod v podatkovno znanost

- področja uporabe v znanosti in tehnologiji

Pridobivanje podatkov

- kako pridobiti podatke in ustvariti podatkovno zbirko
- pregled formatov podatkov
- pridobivanje podatkov iz html datotek in datotek v drugih formatih, branje iz in pisanje v csv, json in xml formate
- pridobivanje podatkov iz spletnih virov, primeri (SURS, FURS, AJPES)
- how to write and run a simple crawler

Predobdelava in skladiščenje podatkov

- kako organizirati podatkovno zbirko
- kako očistiti podatke in odstraniti balast, primeri
- ponovitev osnov podatkovnih baz
- drugi načini skladiščenja podatkov
- strukturirani in nestrukturirani podatki

Kratek uvod v analizo omrežij

- osnove teorije grafov
- kateri podatki so lahko (in kateri ne morejo biti) predstavljeni kot omrežje
- kako analizirati preprosto omrežje, osnovne omrežne mere
- drugi koncepti analize omrežij

Uvod v Velike podatke

- kaj so "Veliki" podatki, kako in zakaj podatki postanejo veliki
- razlike med velikimi in majhnimi podatki
- osnovni algoritmi za rudarjenje velikih podatkov

Na kratko o uporabi v različnih domenah

Content (Syllabus outline):

Introduction to data science

- fields of application in science and technology

Data acquisition

- how to retrieve data and create a dataset
- overview of data formats
- retrieving data from html files and files in other formats, reading from and converting to csv, json and xml formats
- retrieving data from online sources, examples (SURS, FURS, AJPES)
- how to write and run a simple crawler

Data pre-processing and storing

- how to organize the dataset
- how to clean the data and remove artifacts, examples
- brief overview of databases
- other ways of data storing
- structured and unstructured data

Very brief overview of network analysis

- basics of graph theory
- what data can (and cannot) be represented as network
- how to analyze a simple network, basic network measures
- other network analysis concepts

Introduction to Big data

- what is "Big" data, how and why the data became big
- differences with mining "small" data
- basic algorithms for mining big data

Briefly about applications in various domains

- rudarjenje besedil in določanje čustev, primeri
- uporaba v poslovanju, napoved odziva strank
- uporaba v medicini, zagotavljanje varnejšega cepljenja

- text mining and sentiment extractions, examples
- data science in business, predicting customer response
- data science in medicine, making vaccines safer

Temeljni literatura in viri / Readings:

- Hamilton, B. A. (2015): The Field Guide to Data Science. Booz Allen Hamilton Inc. Dostopno prek: <https://wolfpaulus.com/wp-content/uploads/2017/05/field-guide-to-data-science.pdf>
- Leskovec, J., Rajaraman, A., Ullman, J. (2020): Mining Massive Data Sets. 3rd. Ed. Cambridge University Press. Dostopno prek: <http://www.mmds.org/>
- C. Field (2017): The Data Science Handbook. Wiley.
- Levnajić, Z.: Prosojnice iz predavanj pri predmetu Uvod v podatkovne znanosti. Moodle, FIS.

Cilji in kompetence:

Splošne kompetence:

- Splošno razumevanje pomena podatkov
- Sposobnost interpretacije rezultatov podatkovne analize
- Sposobnost uporabe različnih programskih rešitev za analizo podatkov
- Sposobnost iskanja virov in pridobivanja podatkov za izvedbo analize skladno s podanimi zahtevami

Predmetno-specifične kompetence:

- Znanje programiranja za potrebe podatkovne analitike
- Sposobnost pridobivanja in predobdelave podatkovnih zbirk
- Poznavanje metod podatkovne analitike v teoriji in praksi
- Poznavanje osnov velikih podatkov

Objectives and competences:

General competences:

- General understanding of the meaning of data
- The ability to interpret the results of data analysis
- The ability to use various software solutions for data analysis
- The ability to find sources and obtain data to perform the requested analysis

Subject-specific competences:

- Knowledge of programming for data analysis purposes
- Ability to acquire and pre-process a dataset
- Familiarity with data analysis methods, both in theory and practice
- Basics of Big Data

Predvideni študijski rezultati:

Intended learning outcomes:

Po uspešno opravljenem predmetu, bo študent imel:

- dobro razumevanje pridobivanja in organiziranja podatkovnih zbirk
- dobro poznavanje metod podatkovne analitike, vključno z osnovami statističnih in metod analize omrežij
- osnovno poznavanje analize velikih podatkov
- znanje programirana za potrebe podatkovne analitike

Upon successfully completing this course a student will have:

- good understanding of how to acquire and organize a dataset of interest
- knowledge of various data analysis methods, including statistics and basic network analysis
- basics of Big Data analysis
- basic programming skills for data analytics

Metode poučevanja in učenja:

Learning and teaching methods:

Predavanja z aktivno udeležbo študentov. Praktične vaje, kjer študentje uporabljajo metode naučene na predavanjih, analizirajo podatke in razvijajo preproste programe.

Lectures with active participation of students. Practical hands-on exercises, where students use methods learnt at lectures, analyze the data and develop elementary programming codes themselves.

Delež (v %)

/

Načini ocenjevanja:

Assessment:

Weight (in %)

<p>Način (pisni izpit, ustno izpraševanje, naloge, projekt)</p> <ul style="list-style-type: none"> • domače naloge, kjer študentje rešijo zaokroženo celoto nalog (večinoma v obliki pisanja programske kode) • vsak študent individualno pripravi zaključni projekt 	<p>50 %</p> <p>50 %</p>	<p>Type (examination, oral, coursework, project):</p> <ul style="list-style-type: none"> • homeworks where students complete sets of exercises (mostly writing programming codes). • each student shall have his/her individual final project
--	----------------------------	---

Reference nosilca / Lecturer's references:

- K. Ban, M. Perc, Z. Levnajić, Robust clustering of languages across Wikipedia growth, *Journal of the Royal Society Open Science* 4, 171217, 2017.
- I. Tokuda, Z. Levnajić, K. Ishimura, A practical method for estimating coupling functions in complex dynamical systems, *Philosophical Transactions of the Royal Society A* 377, 20190015, 2019.
- A. Zorko, M. Frühwirth, N. Goswami, M. Moser, Z. Levnajić, Heart Rhythm Analyzed via Shapelets Distinguishes Sleep From Awake, *Frontiers in Physiology* 10, 1554, 2020.
- M. Grau Leguía, Z. Levnajić, L. Todorovski, B. Ženko, Reconstructing dynamical networks via feature ranking, *Chaos* 29, 093107, 2019.
- A. Guazzini, F. Stefanelli, E. Imbimbo, D. Vilone, F. Bagnoli, Z. Levnajić, Humans best judge how much to cooperate when facing hard problems in large groups, *Scientific Reports* 9, 5497, 2019.