Predmet:	Analiza velikih količin podatkov
Course title:	Big Data Analysis

Študijski program in stopnja Study programme and level	Študijska smer Study field	Letnik Academic year	Semester Semester
Podatkovne znanosti, magistrski študijski program druge stopnje	-	Drugi	Tretji
The second cycle masters study programme Data Sciences	-	Second	Third

Vrsta predmeta / Course type Obvezni / Obligatory

Univerzitetna koda predmeta / University course code: 2-PZ-MAG-AVKP-2024-02-05

Predavanja Lectures	Seminar Seminar	Vaje Tutorial	Klinične vaje work	Druge oblike študija	Samost. delo Individ. work	ECTS
30	-	20	-	-	90	5

Nosilec predmeta / Lecturer: izr. prof. dr. Biljana Mileva Boshkoska, izr. prof. dr. Zoran Levnajić

Jeziki / **Languages:**	**Predavanja /** **Lectures:**	Slovenski, angleški / Slovene, English
	Vaje / Tutorial:	Slovenski, angleški / Slovene, English

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

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

Prerequisits:

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

Vsebina:

- *Uvod v analizo velikih količin podatkov*: kaj so to veliki podatki, kje jih najdemo, kako jih shranimo?
- *5 »V« velikih podatkov*
- *Grafično predstavljanje velikih količin podatkov*: kateri diagrami so primerni za prikazovanje velikih količin podatkov;

Content (Syllabus outline):

- *Introduction to the big data analysis*: what is big data, where we find it, how to store it?
- *The 5 "V" of Big data*
- *Visualizations of big data*: which diagrams are suitable for representing big data.
- Softwares for storage, retrieval and modelling of Big data (for example NoSQL)

- *Orodja za shranjevanje, dostop in modeliranje velikih podatkov (npr. NoSQL)*
- *Iskanje podobnih enot*: iskanje med najbližnjimi sosedi, povzemanje podatkov z ohranjanjem podobnosti, lokalno občutljive funkcije in razdalje;
- *Podatkovni tokovi*: podatkovni modeli za podatkovne tokove; vzorčenje podatkov; filtriranje podatkov v tokovih; štetje različnih enot v tokovih;
- *Pogosti podatki*: štetje najpogostejših podatkov: ohranjanje pogostih podatkov v glavnem spominu;
- *Metode nadzorovanega in nenadzorovanega učenja prilagojene za velike količine podatkov;*
- Uporaba velikih podatkov v raznih domenah znanosti in podjetništva

- *Search for similar items*: near neighbour search, similarity preserving summaries of sets, locality sensitive functions and distances;
- *Data streams*: the stream data models; sampling data in a stream; filtering streams; counting distinct elements in a stream;
- *Frequent itemsets*: counting the frequent items in a stream, handling larger datasets in the main memory;
- *Supervised and unsupervised learning methods adapted for Big data;*
- Applications and the usage of Big data approaches in various domains of science and business.

Temeljni literatura in viri / Readings:

- Leskovec, Jure, Rajaraman, Anand in Ullman, Jeffrey David (2020): *Mining of Massive Datasets. (2020)* New York: Cambridge University Press.
- Hastie, Trevor, Tibshirani Robert in Friedmanm Jerome (2009): *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.
- Minelli, Michael, Chambers, Michele in Dhiraj, Ambiga (2013): *Big data, big analytics: emerging business intelligence and analytic trends for today's businesses*. Hoboken, New Jersey: John Wiley & Sons.
- Ishikawa, Hiroshi (2015): *Social Big Data Mining*. CRC Press.

Cilji in kompetence:

Učna enota prispeva k razvoju naslednjih splošnih in predmetno specifičnih kompetenc:
Splošne kompetence:

- sposobnost obvladovanja in pretvorbe realnega problema v obliki lažje predstavljivega modela;
- uporaba ustreznih metodoloških pristopov za izvajanje, koordiniranje in organiziranje raziskav;

Predmetno-specifične kompetence:

- sposobnost sinteze izvirnih idej, konceptov in rešitev določenih problemov iz različnih disciplinarnih področij;
- poznavanje in razumevanje širokega nabora aplikacij

Objectives and competences:

The instructional unit contributes to the following general and subject-specific competences:
General competences:

- the ability to manage and transform a real problem into a simplified model;
- utilization of adequate methodological approaches to conduct, coordination and organisation of research;

Subject-specific competences:
- competence to form original ideas, concepts and solutions for specific problems from different disciplines;
- knowledge and understanding of a wide range of applications of

informacijsko komunikacijske tehnologije v sodobni družbi; • poznavanje konceptov in metodologij za analizo velikih količino podatkov. • Osnovna programerska znanja in koncepti za analizo velikih količin podatkov.	information communication technology in the modern society • knowledge of the concepts and methodologies for the analysis of large amounts of data. • Basic programming concepts and skills for Big data analytics.

Predvideni študijski rezultati:	**Intended learning outcomes:**
Znanje in razumevanje: *Študent/študentka:* • razume posebnosti analize velikih količin podatkov v primerjavi s klasičnimi podatkovnimi analizami • spozna metode, ki so primerne za analize tovrstnih podatkov s uporabosodobne odprtokodne opreme	Knowledge and understanding: *The student:* • understands the specificity of big data analysis compared to classical data analysis • learns methods, designed for big data analysis and state of the art open source softwares

Metode poučevanja in učenja:	**Learning and teaching methods:**
• *predavanja* z aktivno udeležbo študentov (razlaga, diskusija, vprašanja, primeri, reševanje problemov, predstavitve) • *vaje* v računalniški učilnici •	• *lectures* (explanation with discussions, questions, case-studies, presentations) • *tutorials* in the computer classroom

Načini ocenjevanja:	Delež (v %) / Weight (in %)	**Assessment:**
Način (pisni izpit, ustno izpraševanje, naloge, projekt): • samostojno pripravljena in predstavljena seminarska naloga, v kateri študent naredi analizo enega vira velikih količin podatkov	100 %	Type (examination, oral, coursework, project): • student individually prepares and presents a project related to analysis of one source of big data

Reference nosilca / Lecturer's references:

- ZHAO, Guoqing, LIU, Shaofeng, LOPEZ, Carmen, LU, Haiyan, ELGUETA, Sebastian, CHEN, Huilan, MILEVA BOSHKOSKA, Biljana. Blockchain technology in agri-food value chain management : a synthesis of applications, challenges and future research directions. *Computers in industry*, ISSN 0166-3615. [Print ed.], 2019, vol. 109, str. 83-99
- BOŠKOSKI, Pavle, DEBENJAK, Andrej, MILEVA BOSHKOSKA, Biljana. Rayleigh copula for describing impedance data - with application to condition monitoring of proton exchange membrane fuel cells. *European journal of operational research*, ISSN 0377-2217. [Print ed.], 2018, vol. 266, no. 1, str. 269-277

- GRAŠIČ, Valerij, KOS, Andrej, MILEVA BOSHKOSKA, Biljana. Classification of incoming calls for the capital city of Slovenia smart city 112 public safety system using open Internet of Things data. International journal of distributed sensor networks, ISSN 1550-1477. [Online ed.], 2018, vol. 14, no. 9, str. 1-12, ilustr.
- KUNIĆ, Zdravko, ŽENKO, Bernard, BOSHKOSKA, Biljana Mileva. FOCUSED-short-term wind speed forecast correction algorithm based on successive nwp forecasts for use in traffic control decision support systems. *Sensors*. 2021, vol. 21, no. 10, str. 3405-1-3405-17. ISSN 1424-8220.
ANDONOVIKJ, Viktor, BOŠKOSKI, Pavle, DŽEROSKI, Sašo, BOSHKOSKA, Biljana Mileva. Survival analysis as semi-supervised multi-target regression for time-to-employment prediction using oblique predictive clustering trees. *Expert systems with applications*. [Online ed.]. Jan. 2024, [article no.] 121246, vol. 235, str. 1-11, ilustr. ISSN 1873-6793.

- M. Grau Leguia, Z. Levnajić, L. Todorovski, B. Ženko, Reconstructing dynamical networks via feature ranking, Chaos 29, 093107, 2019.
- Faggian, F. Ginelli, F. Rosas, Z. Levnajić, Synchronization in time-varying random networks with vanishing connectivity, Scientific Reports 9, 10207, 2019.
- Joksomović, J., Perc, M., Levnajić, Z. Self-organization in Slovenian public spending, Journal of the Royal Society Open Science 10, 221279, 2023.
- Jović, M., Šubelj, L., Golob, T., Makarovič, M., Yasseri, T., Boberić Krstićev, D., Škrbić, S., Levnajić, Z. Terrorist attacks sharpen the binary perception of "Us" vs. "Them", Scientific Reports 13, 12451, 2023.