

**UČNI NAČRT PREDMETA / COURSE SYLLABUS**

**Predmet:** Izbrana poglavja iz analize velepodatkov  
**Course title:** Selected Topics from Big Data Analysis

Študijski program in stopnja Study programme and level	Študijska smer Study field	Letnik Academic year	Semester Semester
Informacijske znanosti, doktorski študijski program tretje stopnje	Računalniške znanosti	Drugi	Tretji ali četrti
Information Sciences, third cycle Doctoral Study Programme	Computer sciences	Second	Third or fourth

**Vrsta predmeta / Course type**

Izbirni / Elective

**Univerzitetna koda predmeta / University course code:**

1-IZ-DR-RZ-IP-IPAV-2024-04-24

Predavanja Lectures	Seminar Seminar	Vaje Tutorial	Klinične vaje work	Druge oblike študija	Samost. delo Individ. work	ECTS
30	-	-	-	-	270	10

**Nosilec predmeta / Lecturer:** Izr. prof. dr. Pavle Boškoski

**Jeziki / Languages:**

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

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

Pogoj za vključitev v delo je vpis v 1. ali 2. letnik študijskega programa.

**Prerequisites:**

A prerequisite for participation is enrollment into the 1st or 2nd year of the study programme.

**Vsebina:**

- Napredna analitična tehnologija in orodja za uporabo HPC za analizo velepodatkov:
  - MapReduce;
  - Apache Hadoop;
  - Posode;
  - Shranjevanje podatkov in NoSQL;
  - Vizualizacija podatkov.
- Algoritemske tehnike za velepodatke in analitiko nestrukturiranih podatkov:

**Content (Syllabus outline):**

- Advanced Analytics Technology and Tools for using HPC to analyse big data:
  - MapReduce;
  - Apache Hadoop;
  - Containers;
  - Data storage and NoSQL;
  - Data visualisation.
- Algorithmic techniques for big data and Analytics of unstructured data:

- Streaming: vzorčenje in skiciranje, rudarjenje;
- Zmanjšanje dimenzij;
- Zunanji pomnilnik in algoritmi poltokovnega pretoka;
- Zasnova algoritma skoraj linearnega časa;
- Testiranje lastnosti;
- Metric Embedding;
- Transformacija redkih podatkov;
- Crowdsourcing.
- Visoka integracija strojnega učenja:
  - Perceptron;
  - Metoda podpornih vektorjev;
  - Nevronske mreže;
  - Globoko učenje.

- Streaming: Sampling and Sketching, mining;
- Dimensionality Reduction;
- External Memory and Semi-streaming Algorithms;
- Near Linear Time Algorithm Design;
- Property Testing;
- Metric Embedding;
- Sparse Transformation;
- Crowdsourcing.
- Large-scale machine learning:
  - Perceptron;
  - Support vector machines;
  - Neural networks;
  - Deep learning.

### Temeljni literatura in viri / Readings:

- Leskovec, J., Rajaraman A., Ullman, J. D. (2022). *Mining of Massive Datasets*, Third Edition. Cambridge university press
- Makino, J. (2021). *Principles of High-Performance Processor Design. For High Performance Computing, Deep Neural Networks and Data Science*. Springer
- EMC Education Services (2015). *Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data*. Willy.
- McElreath, R. (2020). *Statistical rethinking: A Bayesian Course with Examples in R and Stan*. Chapman & Hall/CRC.
- Murphy, K. P. (2012). *Machine learning: A Probabilistic Perspective*. MIT Press.
- Murphy, K. P. (2023). *Probabilistic Machine Learning: Advanced Topics*. MIT Press.
- James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). *An introduction to statistical learning: with Applications in Python*. Springer.

### Cilji in kompetence:

- Učna enota prispeva k razvoju naslednjih splošnih kompetenc:*
- Sposobnost identificiranja danega raziskovalnega problema, njegove analize, ovrednotenja ter oblikovanja možnih rešitev.
  - Ustvarjanje novega znanja, ki pomeni relevanten prispevek k razvoju znanosti.
  - Sposobnost obvladanja standardnih metod, postopkov in procesov raziskovalnega dela na znanstvenem področju študija.
  - Prizadevanje za kakovost znanstveno-raziskovalnega dela skozi avtonomnost, (samo)kritičnost, (samo)refleksivnost in (samo)evalviranje.

### Objectives and competences:

- The module contributes to the following general competencies:*
- Ability to identify a given research problem, analyse it, evaluate it and formulate possible solutions.
  - Ability to create new knowledge, which represents a contribution to science.
  - Ability to master standard methods, procedures and processes of research work in the scientific field of study.
  - Striving for quality in scientific research through autonomy, (self-)criticism, (self)reflexivity and (self-)evaluation.
  - Development of skills and abilities in usage of knowledge in the scientific field of doctoral dissertation.

- Razvoj veščin in spretnosti v uporabi znanja na raziskovalnem področju doktorske disertacije.
- Sposobnost inovativne uporabe in kombiniranja raznih raziskovalnih metod.

In predmetno-specifičnih kompetenc:

- Sposobnost za načrtovanje in konstruiranje rešitve konkretnih raziskovalnih problemov povezanih z analizo velepodatkov.
- Sposobnost pridobivanja, selekcije, kritičnega ocenjevanja in umeščanja novih znanj in zmožnost interpretacije v kontekstu analize velepodatkov.
- Sposobnost oblikovanja rešitve ter njihova implementacija za dane raziskovalne probleme analize velepodatkov.

- Ability to innovatively use and combine diverse research methods.

and subject-specific competences:

- Ability to design and construct solutions for specific research problems and questions related to big data analysis.
- Ability to acquire, select, critically evaluate and situate new knowledge and to interpret it in the context of big data analysis.
- Ability to design and implement original scientific solutions to given research problems in big data analysis.

#### **Predvideni študijski rezultati:**

Znanje in razumevanje:

*Študent/študentka:*

- analizira, primerja ter razume posebnosti analize velepodatkov v primerjavi s klasičnimi podatkovnimi analizami.
- raziskuje in preuči metode, ki so primerne za analize tovrstnih podatkov in z uporabo zelo zmogljivih računalnikov in sodobne odprtokodne opreme izvede analizo na enem primeru velepodatkov.

#### **Intended learning outcomes:**

Knowledge and understanding:

*The student:*

- analyzes, compares and understands the specificity of big data analysis to classical data analysis,
- researches and studies methods, designed for big data analysis and with the use of high-performance computers and state of the art open-source software analyze one instance of big data.

#### **Metode poučevanja in učenja:**

- *predavanja z aktivno udeležbo študentov (razlaga, diskusija, vprašanja, primeri, reševanje problemov, predstavitve),*
- *individualne in skupinske konzultacije (diskusija, dodatna razlaga, obravnava specifičnih vprašanj).*

#### **Learning and teaching methods:**

- *lectures (explanation with discussions, questions, case-studies, presentations),*
- *individual and group consultations (debate, additional explanations, considering specific issues).*

#### **Načini ocenjevanja:**

Delež (v %) /

Weight (in %) /

#### **Assessment:**

Način (pisni izpit, ustno izpraševanje, naloge, projekt): <ul style="list-style-type: none"> <li>projektna raziskovalna naloga</li> </ul>	100	Type (examination, oral, coursework, project): <ul style="list-style-type: none"> <li>project research work</li> </ul>
---	-----	--

**Reference nosilca / Lecturer's references:**

- Brešar, M., & Boškosi, P. (2023). Directional coupling detection through cross-distance vectors. *Physical Review*, 107(4). <https://doi.org/10.1103/physreve.107.044220>
- Boškosi, P., Perne, M., Rameša M., & Boshkoska Mileva, B. (2021). Variational Bayes survival analysis for unemployment modelling. *Knowledge Based Systems*, 229, 107335. <https://doi.org/10.1016/j.knosys.2021.107335>
- Andonovikj, V., Boškosi, P., Džeroski S., & Boshkoska Mileva, B. (2024). Survival analysis as semi-supervised multi-target regression for time-to-employment prediction using oblique predictive clustering trees. *Expert Systems With Applications*, 235, 121246. <https://doi.org/10.1016/j.eswa.2023.121246>
- Andonovikj, V., Boškosi, P., Evkovski, B., Redek, T., & Boshkoska Mileva, B. (2022). Community analysis in Slovenian labour network 2010-2020. *Journal of Decision Systems*, 31(1), 308–318. <https://doi.org/10.1080/12460125.2022.2070944>
- Žnidarič, L., Nusev, G., Morel, B., Mougín, J., Juričić, Đ., & Boškosi, P. (2021). Evaluating uncertainties in electrochemical impedance spectra of solid oxide fuel cells. *Applied energy*, 298, 117101. <https://doi.org/10.1016/j.apenergy.2021.117101>