

COURSE DESCRIPTION

General information		
Lead instructor	Associate Professor Pavle Boškosi Assistant Professor Hrvoje Jerković	
Course name	CDS-15: Selected chapters from big data analysis	
Study programme	Computer and Data Science, third cycle Doctoral Study Programme	
Course status	Optional	
Year	First or Second	
Number of credits and mode of delivery	ECTS student workload coefficient	10
	Number of hours (L+P+S)	30/-/270

Course description	
<i>1.1. Course goals</i>	
<p>The module contributes to the following general competencies:</p> <ul style="list-style-type: none"> - Ability to identify a given research problem, analyze it and offer possible solutions. - Ability to create new knowledge, which represents a contribution to science. - Mastery of research methods, procedures and processes in diverse scientific fields. - Familiarity with the notion of quality and striving for professional quality through autonomy, (self-) criticism, (self-) reflection and (self-) evaluation. <p>and subject-specific competences:</p> <ul style="list-style-type: none"> - Ability to solve specific research problems and questions related to big data analysis. - Development of skills in the use of knowledge in the research field of the doctoral dissertation. - Ability to use innovatively and combine different research methods. - Ability to acquire, select, evaluate and place new knowledge and the ability to interpret in the context of social and other sciences. - Ability to design and implement original scientific solutions to given research problems. 	
<i>1.2. Course enrolment requirements</i>	
A prerequisite for participation is enrollment into the 1st or 2nd year of the study programme.	
<i>1.3. Intended course learning outcomes</i>	
<p>Knowledge and understanding:</p> <p>The student:</p> <ul style="list-style-type: none"> - understands the specificity of big data analysis compared to classical data analysis 	

- learns 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

1.4. Course content

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: 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

1.5. Modes of delivery (mark the appropriate boxes with an X)	<input checked="" type="checkbox"/> lectures	<input checked="" type="checkbox"/> independent work
	<input type="checkbox"/> seminars and workshops	<input type="checkbox"/> multimedia and network
	<input checked="" type="checkbox"/> practicals	<input type="checkbox"/> laboratory
	<input type="checkbox"/> remote learning	<input type="checkbox"/> supervision
	<input type="checkbox"/> field work	<input type="checkbox"/> other _____

1.6. Student obligations

1.7. Monitoring student work (mark the appropriate boxes with an X)

Class attendance		Participation in class		Seminar paper		Experimental work	
Written exam		Oral exam		Essay		Research	
Project		Continuous assessment of knowledge		Student report		Practical work	
Portfolio		Schoolwork		Homework			

1.8. Assessment and evaluation of student work during classes and the final exam

Type (examination, oral, coursework, project):

- Project work, 100%

<i>1.9. Required readings and number of copies relative to the number of students currently taking the course</i>		
<i>Title</i>	<i>Number of copies</i>	<i>Number of students</i>
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.		
<i>1.10. Supplementary readings</i>		
<i>1.11. Methods of quality monitoring that ensure the acquisition of knowledge, skills and competences.</i>		