

Study program:: Artificial Intelligence			
Name of the subject: Large Scale Data Mining			
Teacher(s): Sanja Brdar			
Status of the subject: elective			
Number of ECTS credits: 5			
Conditions: none			
Subject goal			
<ul style="list-style-type: none"> - Introducing the methods for large-scale computational data analysis - Learning programming skills and tools for analysing large-scale data, models' selection and hyperparameter tuning, deploying models - Ability to combine skills from areas such as data storage, distributed systems design, , statistical data analysis, machine learning, graph theory, etc. in order to create value from Big Data 			
Outcome of the subject			
<ul style="list-style-type: none"> - Experience in analysis and processing of massive data sets - Ability to design and implement an analytical solution: choose appropriate storage, algorithms, provide result interpretation and visualisation - Ability to work and solve problems in a variety of data intensive areas 			
Subject content			
<i>Theory</i>			
<ul style="list-style-type: none"> - Data storage and data pre-processing; Features engineering and selection; Integration of data / knowledge / methods; Evaluation metrics; Hyperparameter tuning and model selection, Transfer Learning, Data visualization; 			
<i>Practical learning</i>			
<ul style="list-style-type: none"> - Case studies and applications on heterogeneous data (logs, text, spatio-temporal data, social graphs, etc.) from real-world sources (smart phones, telecom operators, social media, satellite imagery, sensors, genomics) - Implementing solutions in Python with additional packages: Numpy, SciPy, Networkx, Matplotlib, , Scikit-learn, Pandas, PySpark, Keras, PyTorch 			
Expert from the industry will be included into the project assignment realization as an external tutor.			
Literature			
<ol style="list-style-type: none"> 1. Alice Zheng, Evaluating Machine Learning Models, O'Reilly Media, 2015 2. Jason Brownlee, Better Deep Learning: Train. Faster, Reduce Overfitting and Make Better Predictions; Machine Learning Mastery, 2019 3. George Kyriakides, Konstantinos G. Margaritis, Hands-On Ensemble Learning with Python: Build highly optimized ensemble machine learning models using scikit-learn and Keras, Packt Publishing, 2019 4. Mohamed Elgendy, Deep Learning for Vision Systems, Manning Publications 2020 5. Tomasz Drabas, Denny Lee, Learning PySpark, Packt Publish, 2017. 			
Number of active teaching classes	Theoretical teaching: 2		Practical teaching: 2
Method of carrying out the teaching			
Lectures; revisions of the material; active students' participation in problem solving; homework assignments; application of the taught material on real-world examples.			
Evaluation of knowledge (maximum number of points 100)			
Pre-exam obligations	points	Final exam	points
		Oral exam	30
Course project	70		