

General information		
Course leader	Leo Mršić, PhD, Assistant Professor	
Course title	Data preparation and introduction to data visualization	
Study programme		
Course status	Mandatory	
Year	Year 1, semester 1	
Number of credits and mode of teaching delivery	ECTS student workload coefficient	4
	Number of hours (L+E+S)	45 (30 P + 15 V + 0 S)

COURSE DESCRIPTION
1. <i>Course objectives</i>
<p>For the data analysis to have high quality results, it is necessary to make the preparation of the input data. The aim of the course is to demonstrate basic methods of data preparation that includes methods of cleaning, transforming, introverting, normalizing and aggregating data, time series transformation, work with missing values as well as basic data reduction methods such as feature reduction, sample reduction, and discretization. The course objective is to introduce students with basic visualization techniques, explorative data analysis and predictive modeling. It is a necessary theoretical and practical knowledge and skills for all business areas that are characterized by a large amount of data. Besides the technique, students are introduced to various visualization tools, exploratory data analysis tools and predictive modeling tools.</p>
2. <i>Conditions for enrolment in the course</i>
No formal conditions.
3. <i>Expected learning outcomes of the course</i>
<ul style="list-style-type: none"> • L01: Recommend solutions for problems while preparing data. • L02: Choose an adequate method for working with missing data and method of data transformation. • L03: Select the appropriate aggregation functions and methods of time series transformation. • L04: Select an adequate solution for a particular problem in the process of integration, normalization and data discretization. • L05: Apply adequate basic methods of reducing features and patterns. • L06: Explain the impact of new technologies on the process of data preparation • L07: Select appropriate visualization tool, correct visualization errors, and critically interpret choice of analytical patterns and the techniques of analytical interaction used to achieve graphical integrity of given example.

- **L08: Link data from multiple sources for visual analysis, create complex interactive dashboard and argue the choice of granularity and used dashboard elements.**
- **L09: Apply advanced data pre-processing techniques to create features (variables) for predictive models and interpret the results of complex machine learning algorithms and statistical models.**

4. *Course content*

Introduction to data preparation. Data cleaning. Work with missing values. Data Transformation. Sample Reduction. Aggregation of data. Transformation of time series. Data Integration. Normalization of data. Data discretization. Feature Reduction. Practice and Future. Exam preparation.

Introduction to data visualization. The visual apparatus. The most common errors in data visualization. Types of visualization tools. Analytical interactions: comparison, sorting and grouping, aggregation, change the definition of variables. Analysis of the structure, distribution and time series. Forecasting. Analysis of deviations and correlations. Spatial analysis. Dashboards for visualization and exploratory analysis.

5. <i>Teaching delivery modes:</i>	lectures seminars and workshops exercises remote learning field work	independent work multimedia and network laboratory mentoring other
------------------------------------	---	--

6. *Comments*

7. *Student obligations*

STUDENT ATTENDANCE

Class attendance is mandatory in the percentage prescribed by the Studies and examination regulations.

PASSING THE EXAM

The course has defined learning outcomes. In order for a student to pass the course, he/she must achieve a minimum of 50% of the points available for each learning outcome and collect a minimum of 50.01 points out of a possible 100 points per course.

8. *Monitoring^a student work*

Class attendance		Activity during class	10%	Seminar paper		Experimental work	
Written exam	90%	Oral exam		Essay		Research	
Project		Continuous assessment of knowledge		Student report		Practical work	
Portfolio							

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

A grading system based is on a credit accumulation model combined with a defined sub-model, providing a model of the grading method and checking the satisfaction of learning outcomes used in this course.

CONCRETE REVIEW OF EVALUATION METHODS

The maximum number of points that a student can earn in a course is 100. Grades are calculated according to the following criteria table within which the distribution of passing grades in terms of the number of points is applied.

Points	Grade
0,00 - 50,00	(1) unsatisfactory
50,01 - 58,00	(2) sufficient
58,01 - 75,00	(3) good
75,01 - 92,00	(4) very good
92,01 - 100,00	(5) excellent

The method of accumulating points is determined in this course in accordance with the elements of scoring as follows:

Criterion	Maximum points
Written exam	90
Activity during class	10
TOTAL	100

The way of taking the colloquiums, the learning outcomes it covers, as well as the implementation of exams and remedial exams are defined by the "Instructions for attending and taking the course".

10. *Required reading (at the moment of submitting the joint study programme report)*

Salvador García, Julián Luengo, Francisco Herrera: **Data Pre-processing in Data Mining (2016)**

Appavu Balamurugan S., Arockia Christopher A.B.: **Insight into Data Pre-processing: Theory and Practice: Data Mining Perspective (2012)**

Soumen Chakrabarti, Earl Cox, Eibe Frank, Ralf Hartmut Güting, Jiawei Han, Xia Jiang, Micheline Kamber, Sam S. Lightstone: **Data Mining: Know It All (2009)**

Few, S.: **Now You See It: Simple Visualization Techniques for Quantitative Analysis**, Analytics Press, 2009.

Tufte, E.: **Visual Display of Quantitative Information**, Graphics Pr, 2001.

11. *Additional reading (at the moment of submitting the joint study programme report)*

12. *Number of copies of required reading in relation to the number of students who currently attend a course*

Title	Number of copies	Number of students
-------	------------------	--------------------

13. <i>Methods of quality monitoring that ensure the acquisition of knowledge, skills and competencies.</i>		
Monitoring the fulfilment of the desired learning outcomes is an important element of assessment because learning outcomes are the "guarantees" that the school gives to students, but also to employers and the wider community. Learning outcomes represent the minimum threshold that each student must achieve in order to pass the course. For a passing grade, the student must satisfy all the learning outcomes with the demonstrated knowledge, which corresponds to 50% of the points achieved for each learning outcome. The method of scoring based on learning outcomes is presented in the document "Instructions for attending and taking the course".		