SYLLABUS

Name: ***Data Mining in Practice (MakAu-DS>SM2DMP19)***
Name in Polish:
Name in English: ***Data Mining in Practice***

Information on course:

Course offered by department: Faculty of Automatic Control, Electronics and Computer Science
Course for department: Silesian University of Technology

Default type of course examination report:

ZAL

Language:

English

Course homepage:

https://platforma2.polsl.pl/rau2/course/view.php?id=925

Short description:

The aim of the course is to familiarise students with the methodology of implementing the process of data mining, in particular complex-structure data. Specific use cases of data analysis will be presented including the strengths and weaknesses of individual analytical methods. Selected (main) analysis platforms will be discussed and demonstrated.

Description:

Lecture:
1. Cross Industry Standard Process for Data Mining data mining methodology.
2. Data preparation and cleaning (feature selection, feature extraction, missing values, unbalanced data)
3. Analytical model developing, selection and evaluation (classification, regression)
4. Model deployment and maintenance (analytical model deployment, batch mode, incremental mode, concept drift)
5. Social network analysis I (network characteristics and measures, network extraction)
6. Social network analysis II (network visualisation, community identification)

Laboratory classes:
1. Data mining - use cases (Altair AI Studio,, data preparation and cleaning).
2. Data mining – use cases (introduction to deep learning, classification).
3. Social Network analysis – use cases (Pajek, Gephi – network transformation, identification of influential network actors, community identification, network visualisation)

Number of ECTS credits: 2
Total number of hours: 60 (30 contact hours / 30 student’s own work hours)
Lecture: 15h
Laboratory: 15h

Student's own work: preparation to pass the lecture, preparation to the laboratory

Bibliography:

1. Han J., Kamber M., Pei J.: Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers, 2017
2. Witten I. H., Frank E., Hail M.A: Data Mining. Practical Machine Learning Tools and Techniques. Fourth Edition. Theory and its applications, Morgan Kaufmann /Elsevier 2017.
3. Bifet A., Gavaldà R., Holmes G. and Pfahringer B. Machine Learning for Data Streams with Practical Examples in MOA, MIT Press, 2018
4. De Nooy W., Mrvar A. and Batagelj V. Exploratory social network analysis with Pajek: Revised and expanded edition for updated software. Cambridge University Press, 2018
5. Hastie T., Tibishirani R., Freidman J., The Elements of Statistical Learning. Data mining, Inference and Prediction. Springer, 2009
6. McKinney W: Python for Data Analysis. O’Reilly, 2022.
7. Bengio, Y.: Deep Learning. MIT Press, 2016.

Learning outcomes:

Course-specific learning outcomes: at the completion of the course, student:
1. Knows data mining methodology compliant with CRISP DM standard (test, final test) K2A_W05, K2A_W07)
2. Knows data modelling methods adequate to the type of analytical task. The student is able to assess the quality of the obtained models and choose the most appropriate for the particular task (test, final test) K2A_W05; K2A_W07; K2A_W17
3. Knows basic methods of social network analysis (test, final test) K2A_W05; K2A_W07
4. Can apply the results of analysis in the business process (laboratory report) K2A_U01; K2A_U09
5. Can define the methodology for monitoring the quality of analytical models and their modifications during operation (laboratory report) K2A_U01; K2A_U09
6. Can analyze the social network with the selected tool, perform network transformation, indicating important nodes in the network, identifying communities and applying network visualization supporting the analysis (laboratory report) K2A_U01; K2A_U09

Assessment methods and assessment criteria:

DMiP course consists of two components: lecture and laboratory. According to SUT regulation, lecture attendance is optional (however highly recommended), whereas laboratory exercises are obligatory.

Lecture
Written assessment in the form of test containing open or multiple-choice questions
The criterion for passing: minimum 50% of correct answers

Laboratory
Positive assessment of all laboratory reports

The final grade is the grade obtained from the final test.

Making up missed laboratory work is possible on dates agreed upon with the teacher.

The syllabus is valid from academic year 2024/25 and its content cannot be changed during the semester.

Course credits in various terms:

<without a specific program>

Type of credits	Number	First term	Last term
European Credit Transfer System (ECTS)	2	2020/2021-Z	