

Submodule		TM abbreviation
Foundations of Machine Learning for Computer Science: A Practical Journey with Orange Data Mining and Python		Machine Learning for Computer Science
Responsible person	Faculty	
Dr Markus Westner	Computer Science	
Teacher / Lecturer	Frequency of supply	
Dr Luis Miralles		
Teaching form		

Semester of study according to curriculum	Teaching scope [SWS or UE]	Teaching language	Work effort [ECTS credits]
2	4	English	5

Time commitment:

Classroom study	Self-study

Study and examination performance

There will be a project as well as a final exam for evaluating the performance of the students.

Contents

Data mining refers to the process of deploying advanced analytical solutions throughout an organization, from initial planning to final implementation. This module will guide students through a typical life cycle, such as the CRISP-DM model, and examine each stage in detail, including the tasks and technologies involved.

This module covers a variety of data discovery techniques and algorithms that can be used to identify patterns within the data. The main goal of the students in this module is to give an overview of the different steps in data mining such as business and data understanding, data preparation, modelling, evaluating, and deployment. We use both drag-and-drop modules software and some basic programming commands for creating basic pipelines.

Learning objectives: Professional competence

After successful completion of the submodule, students are able to,

1. Explain the role of data mining in an organisation through the different life-cycle stages.
2. Compare and contrast the main data mining techniques in supervised, unsupervised, and reinforcement learning.
3. Assess, combine, and construct data mining pipelines using visual software (e.g., Weka, Orange or similar) or basic coding instructions (e.g., R, Python, or similar).
4. Apply and evaluate various data mining techniques (prediction, clustering, feature selection...) for different problems.
5. Extract and communicate insights from the discovered patterns using data mining solutions.

Learning objectives: Personal competence

After successful completion of the submodule, students are able to

1. Master complex dataset analysis using machine learning for improved problem-solving.
2. Extract and communicate key insights from machine learning patterns effectively.
3. Excel in using programming languages and platforms (e.g., Python and Orange DM) for machine learning pipelines.
4. Apply ethical considerations in machine learning, addressing privacy and bias issues.
5. Acquire project management skills for the entire machine learning life cycle.

Teaching media

The teaching media for the data mining module includes traditional lectures for theoretical foundations, interactive tutorials for hands-on experience, and computer laboratory sessions where students can apply data mining techniques using relevant software tools or programming languages. Visual software, such as Weka or Orange, as well as Python will be utilized for graphical representation of concepts. Additionally, a variety of reading materials, including websites, white papers, research papers, and reports, are provided to supplement the learning process. The combination of these teaching methods aims to offer a comprehensive understanding of data mining concepts, practical application, and ethical considerations within the context of machine learning.

Literature

- Han, J., Pei, J., & Tong, H. (2022). *Data mining: concepts and techniques*. Morgan kaufmann.
- Linoff, G. S., & Berry, M. J. (2011). *Data mining techniques: for marketing, sales, and customer relationship management*. John Wiley & Sons.
- Müller, Andreas C., and Sarah Guido. *Introduction to machine learning with Python: a guide for data scientists*. " O'Reilly Media, Inc.", 2016.
- Alpaydin, Ethem. *Introduction to machine learning*. MIT press, 2020.

The numbers in brackets indicate the levels to be reached: 1 - know, - 2can, - 3understand and apply