Linnæus University

Dr: 2022/29-3.1.2.2

Course syllabus

Faculty of Technology

Department of Computer Science and Media Technology

1DV515 Datautvinning i praktiken, 7.5 credits
Data Mining in Practice

Main field of study
Computer Science

Subject Group
Informatics/Computer and Systems Sciences

Level of classification
First Level

Progression
G1N

Date of Ratification
Approved by Faculty of Technology 2022-01-10
The course syllabus is valid from autumn semester 2022

Prerequisites
General entry requirements for university studies.

Objectives

After completing the course, the student shall be able to:
Knowledge and understanding

- A.1 Describe fundamental concepts and principles for data mining, e.g., data preprocessing, distance measures, and clustering.
- A.2 Explain the most common/typical data mining algorithms and techniques for extracting insights from data.
- A.3 Identify common different categories of data sets based on the contents of the data.

Skills and abilities

- B.1 Given a raw data set: preprocess, clean, and organize the data to prepare it for a data mining pipeline.
- B.2 Given a data mining-related problem: (a) Apply the most common/typical data mining algorithms and techniques for extracting insights from data; and (b) Present the results from data mining algorithms in the appropriate way, using

fundamental visualization concepts.
- B.3 Use suitable (programming- or non-programming-based) tools to implement basic data mining pipelines.

Judgement and approach

- C.1 Critically reflect on data mining from a societal perspective, with respect to ethics and value, as well as considering established knowledge research and state of the art.
- C.2 Make well-grounded decisions to identify which data mining tasks and algorithms/techniques can/should be applied given different data mining requirements and different categories of data sets.

Content

The course introduces basic data mining in an applied way, by focusing on constrained real-world problems (with known scopes) and fundamental principles that can be directly applied with or without programming knowledge. Some examples of areas of applications are (but not limited to): social sciences (digital humanities, learning analytics, etc.), healthcare and life sciences (computational biology, digital health, etc.), forestry and wood technology, and any other areas where large data sets are available and need to be investigated. Some examples of algorithms and techniques that are covered by the course are: search engines, network analysis and community detection, clustering, text mining, and basic visualization.
The following is a summary of the topics that are covered by the course:

- Introduction to data mining.
- Data preprocessing: cleaning, filtering, and structuring.
- How to find similar things, e.g., documents and images.
- Network analysis, e.g., community detection and centrality measures.
- Clustering algorithms, e.g., K-Means and Hierarchical Clustering.
- Dimensionality reduction, e.g., PCA and MDS.
- Extraction of information from unstructured text.
- Data mining from a societal perspective with respect to, e.g., ethical questions, business value, and health.
- Tools and software libraries for data mining.

Type of Instruction

The instruction consists of lectures, practical tutorials, and supervised laboratory sessions. The course may also include guest lectures where representatives from academia and industry discuss how and why they use data mining in their work.

Examination

The examination of the course is divided as follows:

Code	Designation	Grade	Credits
2201	Assignments	AF	2,00
2202	Programming Project	AF	2,50
2203	Oral Exam	AF	2,00
2204	Reflection report	AF	1,00

The course is assessed with the grades A, B, C, D, E, Fx or F.

Grade A constitutes the highest grade on the scale and the remaining grades follow in descending order where the grade E is the lowest grade on the scale that will result in a pass. A grade F means that the student’s performance is assessed as a fail.
Assessment of student performance is made through standalone assignments, one course-long practical project, an oral exam, and a written reflection report. Repeat examination is offered in accordance with Local regulations for courses and examinations at the first- and second-cycle level at Linnaeus University.
To pass the course, a grade E or higher is required for all parts. The final grade is decided from: Practical Project (40%), Practical assignments (30%), Oral exam (20%), and Reflection report (10%).
If the university has decided that a student is entitled to special pedagogical support due to a disability, the examiner has the right to give a customized exam or to have the student conduct the exam in an alternative way.

Objectives achievement

The examination elements are linked to the course objectives in the following ways:

Goal	2201	2202	2203	2204
A.1	☑			
A.2	☑		☑	
A.3	☑			
B.1	☑	☑		
B.2	☑	☑		
B.3	☑	☑		
C.1			☑	☑
C.2	☑	☑	☑	☑

Course Evaluation

During the course or in close connection to the course, a course evaluation is to be carried out. The result and analysis of the course evaluation are to be communicated to the students who have taken the course and to the students who are to participate in the course the next time it is offered. The course evaluation is carried out anonymously.

Other

Grade criteria for the A–F scale are communicated to the student through a special document. The student is to be informed about the grade criteria for the course by the start of the course at the latest.

The course is conducted in such a way that the course participants' experiences and knowledge are made visible and developed. This means, for example, that we have an inclusive approach and strive for no one to feel excluded. This can be expressed in different ways in a course, for example by striving to be gender neutral in language and examples.

Required Reading and Additional Study Material

Mandatory literature:
Leskovec, Jure, Rajaraman, Anand, and Ullman, Jeffrey D., Mining of Massive Datasets, Cambridge University Press, latest edition.Pages: 100 of 511.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). New York: springer.Pages: 50 of 426

Zaki, M. J., & Meira Jr., W. (2014). Data mining and analysis: fundamental concepts and algorithms. Cambridge University Press.Pages: 50 of 604

Compendium of provided scientific articles. Pages: 60