Linnæus University

Dnr: LNU-2023/4320

Course syllabus

Faculty of Technology

Department of Computer Science and Media Technology

4DV510 Datautvinning, 5 högskolepoäng
4DV510 Data Mining, 5 credits

Main field of study
Computer Science

Subject Group
Informatics/Computer and Systems Sciences

Level of classification
Second Level

Progression
A1F

Date of Ratification
Approved 2019-10-28
Revised 2023-11-29 by Faculty of Technology. Assessment methods are revised.
The course syllabus is valid from autumn semester 2024

Prerequisites

- 90 credits in Computer Science (including a degree project at Bachelor level).
- 5 hp machine learning on advanced level (e.g. 4DV660 or equivalent).
- English B/English 6 or the equivalent.

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., distance measures and clustering, and
- A.2 explain the state of the art of data mining with respect to techniques and applications.

Skills and abilities

- B.1 Given a problem and a data set or stream, plan and design a data mining pipeline with respect to what methods should be used for, e.g., dimensionality reduction, distance measure, and clustering, as well as reason about the resulting

data quality,
- B.2 implement some of the algorithms, e.g., MDS and DBSCAN clustering in an efficient way, and
- B.3 use suitable tools to extract data from a given (unstructured) data set or stream.

Judgement and approach

- C.1 Reflect on data mining from a societal perspective with respect to ethics and value, and
- C.2 reason about data quality and what effect it has on the extracted data, as well as how the data quality of the input data can be improved and what the value of such an improvement would be.

Content

The course provides an introduction to data mining and its applications, for example search engines, recommender systems, and text mining. The following topics are covered:

- Introduction to data mining.
- The relationship between data mining and machine learning.
- How to find similar things, e.g., documents and images.
- How to extract structured data from streams.
- Link analysis, e.g., PageRank and HITS.
- Clustering algorithms, e.g., EM, Affinity Propagation.
- Dimensionality reduction, e.g., t-SNE, Self-Organizing Maps.
- Extraction of information from text.
- Applications on the web, e.g., recommendation systems, social network analysis.
- Data mining from a societal perspective with respect to, e.g., ethical questions, business value, and health.
- Data quality.
- Tools and software libraries for data mining.

Type of Instruction

The instruction consists of lectures, seminars, and teacher-supervised laboratory sessions. The course also includes 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
2401	Programming Project	AF	2,00
2402	Oral exam	AF	1,00
2403	Programming assignments	AF	1,00
2404	Reflection report	U/G	1,00

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

The 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. The grade F means that the student’s performance is assessed as fail (i.e. received the grade F).

Assessment of student performance is made through theoretical assignments, programming assignments, and a written exam.

To pass the course, grade E or higher is required for all parts. The final grade is decided from: Programming Project (40%), Oral exam (20%), Programming assignments (20%), and Reflection report (20%)

Resit examination is offered in accordance with Linnaeus University’s Local regulations for courses and examination at the first- and second-cycle levels.
In the event that a student with a disability is entitled to special study support, the examiner will decide on adapted or alternative examination arrangements.

Objectives achievement

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

Goal	2401	2402	2403	2404
A.1		☑		
A.2		☑		☑
B.1	☑	☑		
B.2	☑		☑	
B.3	☑		☑	
C.1				☑
C.2		☑		

Course Evaluation

A course evaluation should be conducted during the course or in connection with its conclusion. The results and analysis of the completed course evaluation should be promptly communicated to students who have completed the course. Students participating in the next course instance should be informed of the results of the previous course evaluation and any improvements that have been made, no later than at the start of the course.

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 using the gender neutral example.

Required Reading and Additional Study Material

- Leskovec, Jure, Rajaraman, Anand, and Ullman, Jeffrey D., *Mining of Massive Datasets*, Cambridge University Press, latest edition. Pages: 400 of 511.
- Compendium of scientific articles.