



ID2222 Data Mining 7.5 credits

Datautvinning

This is a translation of the Swedish, legally binding, course syllabus.

If the course is discontinued, students may request to be examined during the following two academic years

Establishment

Course syllabus for ID2222 valid from Spring 2019

Grading scale

A, B, C, D, E, FX, F

Education cycle

Second cycle

Main field of study

Computer Science and Engineering

Language of instruction

The language of instruction is specified in the course offering information in the course catalogue.

Intended learning outcomes

The course studies fundamentals of data mining, data stream processing, and machine learning algorithms for analyzing very large amounts of data. We will use big data processing platforms, such as MapReduce, Spark and Apache Flink, for implementing parallel algorithms, as well as computation systems for data stream processing, such as Storm and InfoSphere.

After this course, students will be able to mine different types of data, e.g., high-dimensional data, graph data, and infinite/never-ending data (data streams); as well as to program and build data-mining applications. They are also expected to know how to solve problems in real-world applications, e.g., recommender systems, association rules, link analysis, and duplicate detection. Moreover, they will master various mathematical techniques, e.g., linear algebra, optimization, and dynamic programming.

Course contents

- Introduction to Data Mining
- Frequent Itemsets
- Finding Similar Items
- Clustering
- Recommendation Systems
- Mining Data Streams
- Dimensionality Reduction
- Large-Scale Machine Learning

Specific prerequisites

Course literature

The contents of the course are derived from the following two textbooks:

A. Rajaraman and J. D. Ullman, Mining of massive datasets. Cambridge University Press, 2012 (alternative: J. Han, M. Kamber, J. Pei, Data Mining: Concepts and Techniques, 3-rd Ed., Morgan Kaufmann, 2012)

Examination

- LAB1 - Programming Assignments, 3.0 credits, grading scale: P, F
- TEN1 - Examination, 4.5 credits, grading scale: A, B, C, D, E, FX, F

Based on recommendation from KTH's coordinator for disabilities, the examiner will decide how to adapt an examination for students with documented disability.

The examiner may apply another examination format when re-examining individual students.

Written examination. Laboratory tasks.

Ethical approach

- All members of a group are responsible for the group's work.
- In any assessment, every student shall honestly disclose any help received and sources used.
- In an oral assessment, every student shall be able to present and answer questions about the entire assignment and solution.