FID3016 Data Mining 7.5 credits

Data mining

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

Establishment

Grading scale

G

Education cycle

Third cycle

Specific prerequisites

Recommended prerequisites:
Acquaintance with concepts and terminology associated with statistics, database systems, and machine learning; a course on data structures, algorithms, and discrete math (such as ID1020 Algorithms and Data Structures); a course in software systems, software engineering, and programming languages; a course on processing, storing and analyzing massive data (such as ID2221 Data-Intensive Computing).

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 pro-

cessing 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. The course also considers current research topics in data mining with a focus on mining of very large amounts of data.

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. They will master various mathematical techniques, e.g., linear algebra, optimization, and dynamic programming. Moreover, students should be able to describe and apply current research trends in data mining (including methods, algorithms, language support and tools).

Course contents

The course contents includes:
• Introduction to Data Mining
• Frequent Itemsets
• Finding Similar Items
• Clustering
• Recommendation Systems
• Mining Data Streams
• Dimensionality Reduction
• Large-Scale Machine Learning
• Overview of the current research in data mining and its connection to other relevant research areas.

Course literature

The contents of the course are derived from the following textbook:
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

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.

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

Other requirements for final grade

Approved written examination, approved assignments, and approved application of current research (in the form of using it for a research paper, report, or project, etc).

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.