Institutionen för data- och systemvetenskap (DSV)

Stockholms universitet

Kursplan

för kurs på forskarnivå

Data Mining (DAMI) **7.5 Högskolepoäng**

Data Mining (DAMI) **7.5 ECTS credits**

Kurskod:	ML3FU20
Gäller från:	HT 2019
Institution	Institutionen för data- och systemvetenskap

Förkunskapskrav och andra villkor för tillträde till kursen

Only PhD students can participate in the course.

Lärandemål

Knowledge and understanding:

After having completed the course, the student will

- have knowledge of the basic concepts in data mining

- be familiar with basic data mining techniques and algorithms

- be familiar with basic R functions in the area of machine learning

Abilities and skills:

After having completed the course, the student will be able to:

- represent a data set in a form that will be useful for data mining

- know the basic theories behind several data mining techniques

- know how to apply data mining techniques and tools (such as R)

- evaluate the performance of different data mining algorithms

- plan and with adequate methods execute data mining projects within given time frames

- present and discuss results and conclusions

Judgements and values:

After having taken the course, the student will be able to

- critically select appropriate representations, algorithms and tools for a given data mining task

- critically reflect over methodological aspects of a proposed data mining project

Innehåll

Lectures: 13x2 hours

- Lectures will cover algorithmic methods of data science, machine learning techniques, and practical applications

Homework assignments: 3

There will be three written assignments, covering the theoretical aspects of the course. The assignments will involve practical problems using real datasets and Rstudio.

Written Exam: 1

- The written examination is a 4-hour examination on the concepts and techniques taught in the course.

- To Pass the exam a grade of 80 % or higher should be obtained.

Synthesis Paper:

- Two research papers will be selected by each student and a synthesis paper will be written.

- The synthesis paper will involve identifying the common concepts of the two papers, discussing their main strengths and weaknesses, comparing and contrasting the methods and results presented in these papers, and finally, formulating potential research questions for future research.

Contents:

Data mining, machine learning and knowledge discovery

- Data representation

- Association rules: frequent pattern mining and rule mining algorithms

- Clustering: K-means, K-medoids, X-means, Hierarchical Clustering

- Decision trees and rules

- Linear models, naive Bayes and lazy learning
- Combining models

- Deep learning

- Time Series: similarity matching, summarization, regression

- Ranking: Google Pagerank, HITS, In-degree

- Programming using R and Rstudio

Course book:

Introduction to Machine Learning (third edition)

Publisher: MIT Press Year: 2014

ISBN: 9780262028189

Obligatoriska moment

The course is examined continuously.

Examinationsformer

The course is examined through a written exam (4.5 hp) and three assignments accompanied by short online quizzes (3 hp).

The quizzes (maximum of five) will be made available on iLearn2 right after each lecture and they are going

to be solved at home. Quizzes will be part of the "Assignments" grade and will also offer bonus points to the "Assignments" grade.

The four homework assignments are to be done individually. Each assignment will contain a "programming" part (using R) and some short theoretical part (writing a small report or proving some theorem, etc).

The grading of the assignments, the written exam, as well as the whole course uses the goal related seven level ECTS grading.

To be given the grade "E" or higher in the homework assignments the student will need to have accomplished at least 60/100 points in the total assignment/quiz grade.

To obtain an "E" grade at the final exam, the student needs to have demonstrated adequate knowledge of all the course aims. Higher grades than "E" will be given in relation to the results on the written exam.

a. Decision regarding examination are taken by the course leader appointed by the head of the department.

b. Assessment is done according to the seven-point grading scale:

A = Excellent
B = Very good
C = Good
D = Satisfactory
E = Sufficient
Fx = Insufficient
F = Completely Insufficient

c. Grading criteria of the course is announced in the beginning of the course.
The course is examined continuously. Total amount of credits: 7,5.

d. To receive a final grade for the whole course, grade Passed is required for all the modules.

e. In addition, the student who:

- gets grade E on the final exam will be given an opportunity to do a supplement. This means that the student through this supplement may receive grade E, not higher, on the actual exam. Examiner will inform the students who are eligible to do a supplement when the results for the exam are published. Supplement must be submitted in due time and can only be used for raising the grade in the present exam.

- has received at least grade E cannot raise the grade by a new exam

- has taken the same exam twice with the same examiner without passing has a right to a new examiner, provided that no special reasons against this exist.

Transitional provisions:

When the course no longer is distributed or it has been substantially modified, the following applies:

- Uncompleted examination should at first be replaced with other similar exams according to a specific compensation plan - in cases where no such exams can be provided, the student has the right to be examined according to the syllabus once per semester during a three-semester period, starting from the semester after the last time the course was given.

Arbetsform
Lectures: 14x2 hours

Lab tutorials in R: 6x2 hours

Assignments: 3

Quizzes: 5-7

Online lectures: several tutorials in R