Institutionen för data- och systemvetenskap (DSV)

Stockholms universitet

Kursplan

för kurs på forskarnivå

Data Science: teori och praktik **7.5 Högskolepoäng**

Data Science: theory and practice **7.5 ECTS credits**

Kurskod:	ML3FU24
Gäller från:	HT 2017
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
•obtained knowledge of advanced concepts in data science
•explored the literature of a particular topic in data science
•become familiar with theoretical and algorithmic aspects of data science methods
•implemented several data science algorithms and tested them on practical cases

Abilities and skills:

After having completed the course, the student will be able to
•apply data science methods for solving relevant problems
•identify the appropriate data science techniques and tools for addressing relevant research questions in the area
•evaluate the performance of data science methods on different datasets and data domains
•present and discuss results and findings in data science related research

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 science task
•critically reflect over methodological aspects of a proposed data science topic

Innehåll

Lectures: 12 x 2hours
•Lectures will cover algorithmic methods of data science, machine learning techniques, statistical analysis methods, and practical applications (approx. 8 lectures)
•Lectures will also be given by invited industrial leaders in the area of data science, such as IBM, SCANIA, King, Huawei, and (tentatively) Google US (approx. 4 lectures)

Homework assignments: 4
•There will be four written assignments, covering the theoretical aspects of the course. The assignments will

involve solving mathematical and theoretical problems in data science related to the presented material
•The assignments are to be implemented individually
•The assignments will be part of the final examination

Research paper presentation: 1
•Each student will select a research area and present a published research paper on that area to the rest of the class. Each presentation will take 30 minutes.
•This will be part of the final examination

Project assignment: 1
•The project assignment will focus on combining the theoretical and practical aspects of data science. Each student will choose a particular use-case presented by an industrial partner
•The objective of a use-case will be to solve a concrete problem that is given by the industrial partner by:
oLinking the problem to a research area of data science
operforming a literature review on that research area and related to the problem at hand
oformulating a research question and answering it
oproviding concrete results to the industrial partner
owriting a research paper and submitting it to a relevant conference venue
•The project will be part of the final examination

Contents:
•Introduction to data science
•Decision trees and linear classifiers
•Random forests and boosting
•Support vector machines
•Deep neural networks
•Temporal data mining
•Sequential pattern mining and episode mining
•Statistical methods for evaluating data mining results
•Reinforcement learning
•The concept of Big Data: MapReduce and Hadoop
•Practical applications from IBM, SCANIA, King, Huawei, and Google

Obligatoriska moment

12 (2 hour) lectures:
•theory of data science: 8 lectures by Panagiotis Papapetrou (and in part by members of the data science group)
•practice of data science: 4 lectures by industrial partners: IBM (already confirmed), King (already confirmed), Scania (already confirmed), Huawei (pending confirmation), Google US (pending confirmation)

Additional office hours for discussion of the assignments and research project will be offered on demand.

2 lectures:
•These will be lectures given by the students for presenting their chosen research paper
•Each student lecture will take approximately 30 minutes (depending on the number of registered students)

Examinationsformer

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.

Arbetsform

Lectures: 12 x 2hours
•Lectures will cover algorithmic methods of data science, machine learning techniques, statistical analysis methods, and practical applications (approx. 8 lectures)
•Lectures will also be given by invited industrial leaders in the area of data science, such as IBM, SCANIA, King, Huawei, and (tentatively) Google US (approx. 4 lectures)

Homework assignments: 4
•There will be four written assignments, covering the theoretical aspects of the course. The assignments will involve solving mathematical and theoretical problems in data science related to the presented material
•The assignments are to be implemented individually
•The assignments will be part of the final examination

Research paper presentation: 1
•Each student will select a research area and present a published research paper on that area to the rest of the class. Each presentation will take 30 minutes.
•This will be part of the final examination

Project assignment: 1
•The project assignment will focus on combining the theoretical and practical aspects of data science. Each student will choose a particular use-case presented by an industrial partner
•The objective of a use-case will be to solve a concrete problem that is given by the industrial partner by:
oLinking the problem to a research area of data science
operforming a literature review on that research area and related to the problem at hand
oformulating a research question and answering it
oproviding concrete results to the industrial partner
owriting a research paper and submitting it to a relevant conference venue
•The project will be part of the final examination

Contents:
•Introduction to data science
•Decision trees and linear classifiers
•Random forests and boosting
•Support vector machines
•Deep neural networks
•Temporal data mining
•Sequential pattern mining and episode mining
•Statistical methods for evaluating data mining results
•Reinforcement learning
•The concept of Big Data: MapReduce and Hadoop
•Practical applications from IBM, SCANIA, King, Huawei, and Google