Kursnamn   Introduction to Data Science
Name of course

Omfattning (högskolepoäng)   3
ECTS credits

Tidsperiod
Course period
First 7 weeks of Fall Semester
(1 x 2-hour-long lab/lecture per week and
 1 x 2-hour-long practical computer laboratory per week)

Antal platser   30
Maximum number of participants

Undervisningsspråk   English
Language of instruction

Kursens syfte samt motivering till varför den bör vara fakultetsgemensam (max 150 ord)
Aim of course and motivation as to why it should be considered "multidisciplinary" to the extent that the faculty should allocate extra financing.

Data Science is the study of the generalizable extraction of knowledge from data in a practical and scalable manner. A data scientist is characterized by an integrated skill set spanning mathematics, statistics, machine learning, artificial intelligence, databases and optimization along with a deep understanding of the craft of problem formulation in a particular domain to engineer effective solutions. This course will introduce students from various domains of science, engineering and technology to this rapidly growing field and equip them with some of its basic principles and tools. In particular they will be introduced to basic skills needed to collect, store, extract, transform, load, explore, model, evaluate, tune, test and predict using large structured and unstructured datasets from the real-world. The course will use the latest, open-sourced, fast and general engine for large scale data processing.

Kursinnehåll, kursens upplägg samt examinationsform (max 150 ord)
Contents, study format and form of examination

The course will cover the following contents:
- key concepts in distributed fault-tolerant storage and computing, and working knowledge of a data scientist's toolkit: Shell/Scala/SQL/Python/R, etc.
- practical understanding of the data science process:
  - ingest, extract, transform, load and explore structured and unstructured datasets
  - model, train/fit, validate/select, tune, test and prediction (or estimate) with a practical understanding of the underlying mathematics, numerics and statistics
  - communicate and serve the model's predictions to the clients
- practical applications of predictive models for classification and regression, using case-studies of real datasets

There will be assignments involving computer programming and data analysis. The grade is based on attendance, course participation and successful completion of programming assignments.

Målgrupp/er (specifiera ämnen/inriktningar) samt rekommenderade förkunskaper
Target group/s (specify, if possible, subject/specialization) and recommended background

Students from various branches of science, engineering and technology are suited for this introductory course. They are recommended to have basic knowledge of algorithms and some programming experience (equivalent to completing an introductory course in computer science), and some familiarity with linear algebra, calculus, probability and statistics. These basic concepts will be introduced quickly and one could take the course by putting extra effort on these basics as the course progresses.

Huvudansvarig institution
Department with main responsibility Department of Mathematics

Andra inblandade institutioner (specifiera hur).
Other departments involved (specify how).

Kontaktperson/er (namn, e-postadress)
Contact person (name, e-mail address)

Raazesh Sainudiin, raazesh.sainudiin@math.uu.se

Anmälan om kursdeltagande till
Application from course participants should be sent to

raazesh.sainudiin@math.uu.se

Senast
Not later than

1 week before the course starts