



FID3016 Data mining 7,5 hp

Data Mining

När kurs inte längre ges har student möjlighet att examineras under ytterligare två läsår.

Fastställande

Kursplan för FID3016 gäller från och med HT16

Betygsskala

Utbildningsnivå

Forskarnivå

Särskild behörighet

Rekommenderade förkunskaper:

Kännedom om begrepp och terminologi associerad med statistik, databassystem, och maskininlärning; en kurs om datastrukturer, algoritmer, och diskret matematik (som ID1020 Algoritmer och datastrukturer); en kurs i mjukvarusystem, software engineering, och programmeringsspråk; en kurs om att behandla, lagra and analysera massiva datamängder (som ID2221 Data-Intensive Computing).

Undervisningsspråk

Undervisningsspråk anges i kurstillfällesinformationen i kurs- och programkatalogen.

Lärandemål

Kursen behandlar fundamenta inom data mining, data stream processing, och maskininlärningsalgoritmer för att analysera mycket stora datamängder. Vi använder plattformar för big data, såsom MapReduce, Spark och Apache Flink, för att implementera parallella algorit-

mer, och även beräkningssystem för data stream processing, såsom Storm och InfoSphere. Efter denna kurs kommer studenterna att kunna utföra data mining på olika typer av data, t.ex. data av högre dimension, grafdata, och infinita/icke-avslutade data (data streams); liksom att programmera och bygga tillämpningar inom data-mining. De förväntas också att kunna lösa problem i praktiska tillämpningar, t.ex. recommender-system, associationsregler, länkanalys, och detektion av duplikat. Dessutom kommer de att behärska olika matematiska tekniker t.ex. linjär algebra, optimering, och dynamisk programmering. Dessutom ska studenten kunna beskriva och tillämpa aktuella forsknings trender i data mining (inklusive metoder, algoritmer, språkstöd och verktyg).

Kursinnehåll

Innehåll:

- Introduction till Data Mining
- Frekventa Itemmängder
- Att hitta liknande enheter
- Klustring
- Recommender-system
- Mining Data Streams
- Dimensionalitätsreduktion
- Storskalig maskininlärning
- Översikt av aktuell forskning inom data mining och dess koppling till andra relevanta forskningsområden.

Kurslitteratur

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

Examinator beslutar, baserat på rekommendation från KTH:s handläggare av stöd till studenter med funktionsnedsättning, om eventuell anpassad examination för studenter med dokumenterad, varaktig funktionsnedsättning.

Examinator får medge annan examinationsform vid omexamination av enstaka studenter.

Övriga krav för slutbetyg

Godkänd skriftlig tentamina, godkända inlämningsuppgifter och godkänd tillämpning av nuvarande forskning (till exempel: användning i en forskningsartikel, forskningsrapport, eller forskningsprojekt).

Etiskt förhållningssätt

- Vid grupparbete har alla i gruppen ansvar för gruppens arbete.
- Vid examination ska varje student ärligt redovisa hjälp som erhållits och källor som använts.
- Vid muntlig examination ska varje student kunna redogöra för hela uppgiften och hela lösningen.