



# ID2222 Datautvinning 7,5 hp

## Data Mining

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

## Fastställande

Kursplan för ID2222 gäller från och med VT19

## Betygsskala

A, B, C, D, E, FX, F

## Utbildningsnivå

Avancerad nivå

## Huvudområden

Datalogi och datateknik

## Särskild behörighet

## Undervisningsspråk

Undervisningsspråk anges i kurstillfällesinformationen i kurs- och programkatalogen.

## Lärandemål

Kursen behandlar fundamenta inom datautvinning, bearbetning av dataströmmar, och maskininlärningsalgoritmer för att analysera mycket stora datamängder. Vi använder plattformar för storskaliga datamängder, såsom MapReduce, Spark och Apache Flink, för att

implementera parallella algoritmer, och även beräkningssystem för bearbetning av dataströmmar, såsom Storm och InfoSphere.

Efter denna kurs kommer studenterna att kunna utföra datautvinning på olika typer av data, till exempel data av högre dimension, grafdata och infinita/icke-avslutade data (dataströmmar); liksom att programmera och bygga tillämpningar inom datautvinning. De förväntas också att kunna lösa problem i praktiska tillämpningar, till exempel rekommendationssystem, associationsregler, länkanalys, och detektion av duplikat. Dessutom kommer de att behärska olika matematiska tekniker till exempel linjär algebra, optimering, och dynamisk programmering.

## Kursinnehåll

- Introduktion till datautvinning
- Frekventa Itemmängder
- Att hitta liknande enheter
- Klustering
- Rekommendationssystem
- Datautvinning från strömmar
- Dimensionalitätsreduktion
- Storskalig maskininlärning

## Kurslitteratur

Innehållet i kursen hämtas från dessa två kursböcker:

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

- LAB1 - Programmeringsuppgifter, 3,0 hp, betygsskala: P, F
- TEN1 - Tentamen, 4,5 hp, betygsskala: A, B, C, D, E, FX, F

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.

Skriftlig tentamen. Laborativa uppgifter.

## 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.